Artificial Intelligence 340 (2025) 104289

Contents lists available at ScienceDirect

Artificial Intelligence

journal homepage: www.elsevier.com/locate/artint

Maximum Likelihood Evidential Reasoning

Jian-Bo Yang *, Dong-Ling Xu

Decision and Cognitive Sciences Research Centre, The University of Manchester, UK

ARTICLE INFO

Keywords:
Evidential reasoning
Bayesian inference
Likelihood analysis of imperfect data
Dempster principle
Interpretable machine learning
Explainable artificial intelligence

ABSTRACT

In this paper, we aim at generalising the evidential reasoning (*ER*) rule to establish a new maximum likelihood evidential reasoning (*MAKER*) framework for probabilistic inference from inputs to outputs in a system space, with their relationships characterised by imperfect data. The *MAKER* framework consists of three models: system state model (*SSM*), *evidence* acquisition model (*EAM*) and evidential reasoning model (*ERM*). *SSM* is introduced to describe system output in the form of ordinary probability distribution on singleton states of the system space to model randomness only, or more generally basic probability distribution on singleton states and their subsets, referred to as states for short, to depict both randomness and ambiguity explicitly. *EAM* is established to acquire evidence from a data source as system input in the form of basic probability distribution on the evidential elements of the data source, with each evidential element pointing to a state in the system space. *ERM* is created to combine pieces of acquired evidence, with each represented in the form of basic probability distribution on all the states and the powerset of the system space to facilitate an augmented probabilistic inference process where the trustworthiness of evidence is explicitly modelled alongside its randomness and ambiguity.

Within the *MAKER* framework, the trustworthiness of evidence is defined in terms of its reliability and expected weight to measure the total degree of its support for all states. Interdependence between pairs of evidence is also measured explicitly. A general conjunctive *MAKER* rule and algorithm are then established to infer system output from multiple inputs by combining multiple pieces of evidence that have weights and reliabilities and are dependent on each other in general. Several special *MAKER* rules and algorithms are deduced to facilitate inference in special situations where evidence is exclusive or independent of each other. Specific conditions are identified and proven where the *MAKER* rule reduces to the *ER* rule, Dempster's rule and Bayes' rule. A bi-objective nonlinear pre-emptive minimax optimisation model is built to make use of observed data for optimal learning of evidence weights and reliabilities by maximising the predicted likelihood of the true state for each observation. Two numerical examples are analysed to demonstrate the three constituent models of the *MAKER* framework, the *MAKER* rules and algorithms, and the optimal learning model. A case study for human well-being analysis is provided where data from a panel survey are used to show the potential applications of the *MAKER* framework for probabilistic reasoning and decision making under different types of uncertainty.

* Corresponding author.
E-mail address: jian-bo.yang@manchester.ac.uk (J.-B. Yang).

https://doi.org/10.1016/j.artint.2025.104289
Received 23 November 2023; Received in revised form 6 January 2025; Accepted 6 January 2025
Available online 15 January 2025

1. Introduction

In this paper, we aim to generalise the *ER* rule that was established to enable probabilistic inference by combining multiple pieces of evidence, with each having weight and reliability and all being independent of each other [71]. Since its establishment, persistent efforts have been made to improve and apply the *ER* rule in various fields, including professional services such as healthcare [1,25–28,31,77,87,88], insurance [35,36,45] and finance [38,43,44,51], multiple criteria decision analysis [17,19,20,34,81–85], machine learning [8–10,61,62,78], engineering and manufacturing systems [29,50,52,63], environment and sustainability [48,49,60], energy systems [30,57], transportation systems [7,72,79], project management [32,33,53,86], and social systems [11,42,54,55,75,76], among others.

From the extensive work on improving and applying the *ER* rule, such as those mentioned above, lessons have been learnt, which have inspired researchers to enhance and generalise it for wider and more robust applications. One such key lesson is about the interpretability and transparency of the inference process, essential for applications in many areas, such as professional services, where decision makers (doctors, lawyers, claim handlers, loan underwriters, etc.) would not accept recommendations generated by *AI* *D*ecision *S*upport *S*ystem*s* (*AI-DSSs*) unless it can provide interpretations on how the recommendations are generated step-by-step and why and to what extent they can be trusted [43,44]. In the *ER* rule, weight and reliability are introduced as one of its unique features; however, there is a lack of explicit definitions and procedures to guide how they should be interpreted and acquired from historical data, human knowledge or a mix of them. With surging interests in big data and *AI* from almost every corner of the world more than ever before, it is inevitable that any application of the *ER* rule needs to make use of data, although it is always advisable not to underestimate the importance and value of human knowledge. Another key lesson learnt is also related to data, specifically, how to acquire evidence from data and use it in inference when multiple pieces of evidence tend to be dependent on each other in general, rather than always independent of each other. It is the development of the likelihood analysis method [72] that creates a robust process to acquire evidence from data and makes it possible to model interdependency explicitly among multiple pieces of evidence, to be explored further in this paper.

In fact, transparency, interpretability and dependency are key issues not unique to *AI-DSS*s built upon the *ER* rule but common to any *AI* systems that are also built to learn, reason and decide. For instance, Mittelstadt [39] stated that *AI* systems are frequently thought of as opaque, meaning their performance or logic is thought to be inaccessible or incomprehensible to human observers, and emphasized that explaining the functionality and behaviour of *AI* systems in a meaningful and useful way to people designing, operating, regulating, or affected by their outputs is extremely challenging. Markus et al. [37] argued that *AI* has huge potential to improve the health and well-being of people, but adoption in clinical practice is still limited, and lack of transparency is identified as one of the main barriers to implementation, as clinicians should be confident the *AI* system can be trusted. Balasubramaniam et al. [2] analysed the ethical guidelines and revealed that the importance of transparency is highlighted by almost all of the organizations and explainability is considered as an integral part of transparency. Ribeiro et al. [41] investigated dependency factors in the Dempster-Shafer theory by proposing a new approach based on Bayesian net, Pearson's test, and linear regression for multi-sensor information fusion and applied it in adverse drug reactions, emphasising the importance of addressing the issue of dependency among factors.

To deal with those key issues it is necessary to explain the functionality and behaviour of *AI* systems in a meaningful and useful way [39]. In this paper, we focus our efforts on addressing the issues of transparency, interpretability and dependency related to the *ER* rule and identified from its applications in theoretical, methodological and technical aspects. That is, when evidence is acquired from imperfect data [72], how should weight and reliability be defined and interpreted and how should interdependence among multiple pieces of evidence be measured and taken into account in reasoning and decision making? It is possible to adjust probabilities or beliefs for better data fusion and mitigate possible correlations or dependencies among evidence using a mixed approach [41]. However, we believe that complying with widely recognised scientific principles is a prerequisite for any reasoning approach to be considered intrinsically transparent and interpretable. It is due to this belief that in this paper we will investigate the theoretical, methodological and technical issues in a systematic manner and endeavour to explore a principled approach for establishing a unique *MAKER* framework.

In the *MAKER* framework, evidence weight and reliability are explicitly defined and can be learnt from data, the trustworthiness of evidence or any inferred conclusion is clearly defined in relation to evidence weight and reliability, and interdependence among evidence is also precisely defined and can be calculated from data. The original thinking that leads to the creation of the unique *MAKER* framework not only results from the persistent efforts to improve and apply the *ER* rule but also is inspired by Dempster's pioneer work on generalising Bayesian inference to model both randomness and ambiguity that are common in imperfect data [12,13] and his system view on how ambiguity should be handled [14], as well as Shafer's creative thinking on the nature of evidence as described in the first three chapters of his book [46].

The *MAKER* framework takes a system view in terms of using evidence acquired from data as system inputs to infer conclusions as system outputs in order to support evidence-based decision making with single criterion or multiple criteria under different types of uncertainty [4,16,18,19,24,47,56,58,64,66–68,85] and for probabilistic modelling and inference based on the belief rule base methodology [9,45,59,69,70]. It consists of three models: *SSM, EAM* and *ERM*, where evidence is profiled as basic probability distributions for probabilistic inference and both input and output models as well as the inference processes are described in the format of evidence.

SSM is rooted in Dempster's pioneer work and system view for generalising Bayesian inference [12–14] and is probably the most important of his work and views. The innovative thinking behind it is unambiguously footed on the widely recognised principle of Bayesian inference that is perhaps a perfect inference engine given perfect data with randomness explicitly taken into account [3,21, 22]. *SSM* sets the foundation of the *MAKER* framework and is first introduced to describe system output in the form of ordinary probability distribution on singleton states of a system space to model randomness only, or more generally in the form of basic probability distribution on both singleton states and their subsets of the system space, all referred to as states for short, to depict both randomness and ambiguity [14].

EAM is established to acquire evidence from a data source as system input in the form of basic probability distribution on the evidential elements of the data source, with each evidential element pointing precisely to a state in the system space. Note that every data source is equipped with its own *EAM*, where each of its own evidential elements corresponds to a specific state in the system space. The creation of *EAM* for a data source provides a basis to acquire evidence from the data source and enables the definition and interpretation of weight as a measure of the data source's ability to provide correct judgements for reasoning and decision making.

ERM is created to profile acquired evidence in the form of basic probability distribution on all the states and the powerset of the system space and provide a basis to model the reliability and trustworthiness (or untrustworthiness) of the acquired evidence. The reliability of a piece of evidence measures the extent to which it can represent a correct outcome (conclusion or assessment) in terms of probability distribution on all states. The trustworthiness of evidence is defined in terms of its reliability and expected weight to measure the total degree of its support for all states. It is within *ERM* that probabilistic inference is performed by combining multiple pieces of evidence with randomness, ambiguity and untrustworthiness all duly taken into consideration under the united framework. In the inference process, interdependence among evidence is explicitly taken into account to infer complex system behaviours in a probabilistic manner. An interdependence index among pieces of evidence is defined precisely and calculated from probabilities.

A general conjunctive *MAKER* rule and algorithm are established to infer system output from system inputs by combining multiple pieces of evidence that each have weights and reliability and are dependent on each other in general. The establishment of the *MAKER* rule is underpinned by the widely recognised principles governing probabilistic inference, in particular, Bayesian principle and the likelihood principle [3,5,22,72], as well as Dempster's principle as introduced in this paper. Several special *MAKER* rules and algorithms are deduced to facilitate inference in special cases where multiple pieces of evidence are either exclusive or independent of each other. The concepts of mutual exclusiveness and independence are explicitly defined in terms of basic probability functions within the *MAKER* framework to facilitate probabilistic inference with imperfect data or knowledge. Specific conditions are identified and proven where the *MAKER* rule reduces to the *ER* rule, Dempster's rule and Bayes' rule.

To facilitate the application of the *MAKER* rules and algorithms, a bi-objective nonlinear minimax optimisation model is built to make use of observation data for optimal learning of evidence weights and reliabilities by means of maximising the predicted likelihood of the true state for each observation. The two objectives are prioritised, and a solution method is proposed to solve the optimisation problem in two steps to optimise the two objectives in sequence. The model is non-smooth in its original format and is transformed to an equivalent smooth model, so that it can be solved by the Evolutionary engine or the GRG Nonlinear engine of Excel Solver, or similar optimisation tools in other platforms.

Two illustrative examples are analysed to demonstrate the three basic models of the *MAKER* framework, the *MAKER* rules and algorithms, and the optimal learning model. The first example is related to disease diagnosis via different tests, e.g. saliva test and blood test, and is composed from experiences gained from many research projects in healthcare. It is so designed that typical features of imperfect data are included and analysed. It is used throughout the paper to elaborate the new concepts and formulae of the *MAKER* framework, rules and algorithm as well as the optimal learning model. The second example is dedicated to illustrating the *MAKER* framework in detail by showing the step-by-step calculations of how it can be applied to mimic the process of a jury reaching its verdict by combining jurors' conclusions that are not fully reliable in nature and are assumed to be mutually independent [40]. A case study for human well-being analysis is provided, where the *B*ritish *H*ousehold *P*anel *S*urvey (*BHPS*) data [23,73,74] is used to show why the *MAKER* framework is useful to analyse the data of high quality yet still imperfect due to missing and imbalanced data, and how widely it may be applied for reasoning and decision making under different types of uncertainty. The detailed calculation processes for the examples and case study are recorded in Excel sheets together with Solver models that are provided as the supplementary materials of the paper.

2. The likelihood method for analysis of imperfect data

The type of probabilistic inference investigated in this paper is based on a system view and concerned with how to acquire evidence from data by means of probability distributions as system inputs, and how to generate system output as probability distributions by combining multiple pieces of evidence. The main idea and requirements of such probabilistic inference are explained using an example discussed in this section, in particular how to acquire evidence from data using a likelihood method. This example is used to demonstrate the original thinking, theoretical developments, modelling processes and new algorithms throughout this paper.

2.1. A data-driven disease diagnosis example

The example is concerned with how to use routinely collected data to predict disease given different tests, such as saliva test and blood test, among others. In such inferential model-based prediction, disease outcome is taken as system output, with saliva test and blood test as system inputs. The output variable is labelled as y and takes three categorical values: whether it is true ($y = y_1$), false ($y = y_2$) or unknown ($y = \Theta = \{\text{true}, \text{false}\}$) that a patient has disease. The first input variable is the saliva test result, is labelled by x_1 and takes two categorical values: saliva positive ($x_1 = x_{11}$) or saliva negative ($x_1 = x_{12}$). The second input variable is the blood test result, is labelled by x_2 and takes two categorical values: blood positive ($x_2 = x_{21}$) or blood negative ($x_2 = x_{22}$). To simplify description, the records having the same input and output values are grouped together. The simulated data of 6300 patient records for the example are shown in Table 1.1 and Table 1.2, where complete and incomplete patient records are listed separately to facilitate discussions in the following sections.

The data shown in Table 1.1 and Table 1.2 are imperfect in the sense that they involve different types of uncertainty, that is, randomness, ambiguity, incompleteness, inaccuracy, and insufficiency caused by imbalance or lack of data. The data is random because the same test results can lead to different outputs. For example, patients 1-93 have positive saliva and positive blood test results and have the disease; however, patients 101-120 also have positive saliva and positive blood test results but do not have the disease. Some of the data is ambiguous in that it is unknown whether the following patients have the disease or not: patients 1001-1100, 3101-3200, and 6201-6300. The data is incomplete in that patients 1101-3200 only have saliva test results, whilst patients 3201-6300 only have blood test results. There is also an issue of data insufficiency caused by data imbalance, which affects data quality and is not obvious from Table 1.1 and Table 1.2 but will be revealed by descriptive data analysis later. In addition to the above, one may also be concerned that the test results and the diagnostic outcomes of the collected data can be inaccurate due to, for example: (i) possible errors in the test, diagnosis and data recording processes, (ii) the number of collected records not being large enough for estimation of population behaviours with high trustworthiness, (iii) inadequacy of using only saliva test and blood test to make reliable prediction of the disease, (iv) the relationships between the disease and saliva test or blood test not being adequately captured by the data of Table 1.1 and Table 1.2, among other possible reasons. That is, the veracity of the data can be a concern, or how trustworthy it is to use the test results to predict the disease.

2.2. Representation and descriptive analysis of imperfect data

Conventional probabilistic inference relies on complete and unambiguous data. This means that only the records of patients 1-1000 of Table 1.1 may be used for inference, while the other records with missing or unknown values may be deleted or imputation may be used to estimate missing or unknown values. However, deleting incomplete or ambiguous data can lead to loss of vital information, and imputing data can change the nature of original data with the risk of leading to less reliable data and potentially biased or distorted outcomes [72].

This paper aims to establish the *MAKER* framework to handle imperfect data in modelling and inference without data deletion or imputation. In the *MAKER* framework, the data shown in Table 1.1 and 1.2 is represented as three parts: 1) the complete patient records shown in Table 1.1, where both test results are recorded, 2) all patient records with saliva tests as shown in Table 2, and 3) all patient records with blood tests as shown in Table 3. Note that the records of Table 1.1 are also used for constructing Table 2 and Table 3, to ensure that all records having saliva test (or blood test) are used to estimate the relationship between the disease and saliva test (or blood test). While data analysis with larger datasets should be of better quality, this makes Table 2 and Table 3 dependent on each other.

The 1100 records of Table 1.1 hold information about the interdependency between saliva test and blood test in their joint support for diagnosis of the disease, although such support may not be deemed fully reliable due to the relatively small sample size, among other concerns. To make sense of the data in Table 1.1, a joint contingency table is constructed in Table 4, where H_{SBtT} is an evidential element for the data source of Table 1.1, standing for "both tests point to disease being true", H_{SBtF} is for "both tests point to disease being false", and H_{SBtU} is for "both tests point to disease being unknown". Introduction of evidential elements implies that even if the two tests jointly point to the disease being true (or false or unknown) to certain extents, as shown in Table 4, it does not necessarily mean that the disease is true (or false or unknown) to the same extents due to such factors as the limited sample size of 1100 records, possible errors in gathering and processing the data and data imbalance, among others.

The data in Table 4 is significantly imbalanced. For instance, the ratio of the number of patients without disease (false) over that with disease (true) is 9 (900/100); among the patients with negative saliva test, the ratio of the number with negative blood test over that with positive blood test is more than 32 (873 / 27).

Table 4 is used to explore the interdependence between saliva test and blood test for disease prediction. Table 5 and Table 6 are constructed from Table 2 and Table 3 for acquiring evidence from saliva test or blood test, respectively. In Table 5, H_{StT} is an evidential

Table 1.1
Complete Patient Records.

Record No.	Input variables		Disease Output (y)
	Saliva (x_1)	Blood (x_2)	
1-93	Positive	Positive	True
94-95	Positive	Negative	True
96-99	Negative	Positive	True
100	Negative	Negative	True
101-120	Positive	Positive	False
121-160	Positive	Negative	False
161-180	Negative	Positive	False
181-1000	Negative	Negative	False
1001-1043	Positive	Positive	Unknown
1044-1045	Positive	Negative	Unknown
1046-1048	Negative	Positive	Unknown
1049-1100	Negative	Negative	Unknown

Table 1.2
Incomplete Patient Records.

Record No	Input variables		Disease Output (y)
	Saliva (x_1)	Blood (x_2)	
1101-1280	Positive	–	True
1281-1295	Negative	–	True
1296-1450	Positive	–	False
1451-3100	Negative	–	False
3101-3150	Positive	–	Unknown
3151-3200	Negative	–	Unknown
3201-3450	–	Positive	True
3451-3470	–	Negative	True
3471-3610	–	Positive	False
3611-6200	–	Negative	False
6201-6254	–	Positive	Unknown
6255-6300	–	Negative	Unknown

Table 2
Data Records for Saliva Test.

Record No.	Saliva Test (x_1)	Disease (y)
1-95	Positive	True
96-100	Negative	True
101-160	Positive	False
161-1000	Negative	False
1001-1045	Positive	Unknown
1046-1100	Negative	Unknown
1101-1280	Positive	True
1281-1295	Negative	True
1296-1450	Positive	False
1451-3100	Negative	False
3101-3150	Positive	Unknown
3151-3200	Negative	Unknown

Table 3
Data Records for Blood Test.

Record No.	Blood Test (x_2)	Disease (y)
1-93,96-99	Positive	True
94-95,100	Negative	True
101-120,161-180	Positive	False
121-160, 181-1000	Negative	False
1001-1043, 1046-1048	Positive	Unknown
1044-1045, 1049-1100	Negative	Unknown
3201-3450	Positive	True
3451-3470	Negative	True
3471-3610	Positive	False
3611-6200	Negative	False
6201-6254	Positive	Unknown
6255-6300	Negative	Unknown

Table 4
Contingency Table for Both Saliva and Blood Tests.

Frequency	Positive Saliva		Negative Saliva		Row total
	Positive Blood	Negative Blood	Positive Blood	Negative Blood	
H_{SBtT}	93	2	4	1	100
H_{SBtF}	20	40	20	820	900
H_{SBtU}	43	2	3	52	100
Column total	156	44	27	873	1100

element for the data source of Table 2, standing for "Saliva test points to disease being true", H_{StF} is for "Saliva test points to disease being false", and H_{StU} is for "Saliva test points to disease being unknown (i.e. either H_{StT} or H_{StF})". In Table 6, H_{BtT} is an evidential element for the data source of Table 3, standing for "Blood test points to disease being true", H_{BtF} is for "Blood test points to disease

Table 5
Contingency Table for Saliva Test.

Frequency	Positive Saliva	Negative Saliva	Row total
H_{StT}	275	20	295
H_{StF}	215	2490	2705
H_{StU}	95	105	200
Column total	585	2615	3200

Table 6
Contingency Table for Blood Test.

Frequency	Positive Blood	Negative Blood	Row total
H_{BtT}	347	23	370
H_{BtF}	180	3450	3630
H_{BtU}	100	100	200
Column total	627	3573	4200

being false", and H_{BtU} is for "Blood test points to disease being unknown (i.e. either H_{BtT} or H_{BtF})". Note that the data for saliva test in Table 5 is imbalanced, so is the data for blood test in Table 6, leading to the prior distribution of Table 5 being different from that of Table 6.

2.3. Likelihood analysis of imperfect data for evidence acquisition

In this subsection, the likelihood method for analysing imperfect data [72] is discussed using the above example, for acquisition of evidence and in preparation for the theoretical discussions of the later sections. In Table 4, there is ambiguous data as shown in the 2nd last row. This makes it inappropriate to use conventional probabilistic inference approaches to draw conclusions directly. For example, out of the 1100 patients who have both saliva and blood tests, 93 patients having the disease are positive in both tests. However, it is inappropriate to conclude that the probability of a patient having the disease given both positive saliva and blood tests is precisely 0.5962 (93/156) due to the ambiguity. There are 43 patients who also have both positive tests, but they may or may not have the disease. This makes it difficult to interpret joint probabilities calculated using the data of Table 4 alone.

The data of Table 4 is not sufficient or balanced, leading to questions on whether it is trustworthy to use Table 4 alone for inference. For instance, it is argued that as a general rule of thumb the expected count per cell in a contingency table should be no less than 5 for inference to be valid [6,80], and some other rules of thumb have also been suggested. For example, Yates et al. [80] suggested that "No more than 20% of the expected counts should be less than 5 and all individual expected counts should be greater than or equal to 1. Some expected counts can be <5, provided none <1, and 80% of the expected counts should be equal to or greater than 5". Table 7 is generated by calculating expected counts for all cells of Table 4 using a simple approach of Row Total times Column Total divided by Total, e.g. 14.18 =100*156/1100 for the cell at the H_{SBtT} row and the "Saliva Positive - Blood Positive" column.

In Table 7, there are 4 expected counts < 5, so the general rule of thumb is not satisfied. Yates' rule of thumb is not met either as in Table 7 there are more than 33% (4/12) of expected counts < 5. One may argue that compliance with these rules of thumb may not guarantee the inference to be always valid. On the other hand, if they are violated by data, one has to question how trustworthy the inference could be if it is purely driven by the data.

The above discussions call for new thinking for probabilistic inference with imperfect data. A new framework to be established in this paper will be based on acquisition of evidence as likelihood distribution generated from imperfect datasets. In Table 5, for example, the likelihoods of observing positive saliva test given states H_{StT}, H_{StF} and H_{StU} are calculated as 0.9322 (275/295), 0.0795 (215/2705) and 0.475 (95/200), respectively. By normalising these likelihoods so that they sum to one [72], we get the normalised likelihood of 0.627 (0.9322/(0.9322+0.0795+0.475) for H_{StT}, 0.0535 for H_{StF}, and 0.3195 for H_{StU}, as shown in the second column of Table 8. These normalised likelihoods formulate a likelihood distribution for profiling a positive saliva test result, referred to as a piece of evidence. Similarly, a negative saliva test result is profiled by another likelihood distribution in the last column of Table 8; positive and negative blood test results are profiled as two pieces of evidence as shown in the last two columns of Table 9.

The data of Table 5 and Table 6 is regarded to be of good quality in the sense that expected counts in cells in Table 5 and Table 6 are all much larger than 5. So, evidence acquired from these tables should also be of good quality. The question of interest is then how to

Table 7
Expected Counts for Both Saliva and Blood Tests.

Expected counts	Positive Saliva		Negative Saliva	
	Positive Blood	Negative Blood	Positive Blood	Negative Blood
H_{SBtT}	14.18	4.00	2.45	79.36
H_{SBtF}	127.64	36.00	22.09	714.27
H_{SBtU}	14.18	4.00	2.45	79.36

Table 8
Normalised Likelihood for Saliva.

Likelihood $p_{H_{St\theta},1i}$	Positive Saliva (x_{11})	Negative Saliva (x_{12})
H_{StT}	0.6270	0.0448
H_{StF}	0.0535	0.6083
H_{StU}	0.3195	0.3469

Table 9
Normalised Likelihood for Blood.

Likelihood $p_{H_{Bt\theta},2j}$	Positive Blood (x_{21})	Negative Blood (x_{22})
H_{BtT}	0.6305	0.0411
H_{BtF}	0.0333	0.6283
H_{BtU}	0.3362	0.3306

use evidence acquired from individual tests, as shown in Table 8 and Table 9, for likelihood inference.

By applying the above process to Table 4, the joint likelihoods of observing positive or negative saliva and blood tests given H_{SBtT}, H_{SBtF} and H_{SBtU} can be calculated, and then normalised to construct Table 10. The normalised joint likelihoods provide information about the interrelationship between saliva and blood tests for disease diagnosis, which will be explored in the following sections.

3. Basic models of the *MAKER* framework

From a system point of view, inference is a systematic process of predicting the output of a system from its inputs, in which the interdependence and interaction among the inputs and between the inputs and output need to be taken into account. If an explicit mathematical function between system inputs and output is known *a priori*, inference could be as direct as to use the known function to calculate the output for given inputs; otherwise, data or judgements about system relationships and behaviours should be gathered and described to enable inference.

It is to enable data-driven and knowledge-based inference that the *MAKER* framework is established. In this section, the concepts and constituent components of the *MAKER* framework will be explored, including a system state model to represent system output associated with randomness and ambiguity, an evidence acquisition model to acquire evidence as system inputs from imperfect data, and an evidential reasoning model to enable inference in terms of evidence combination with the weights of variables and the reliability of evidence being explicitly and consistently taken into account together with randomness and ambiguity.

3.1. *System State Model – SSM*

The concept of System State Model (*SSM* for short) was proposed by Dempster following his pioneer work [12,13] and his system view [14]. In *SSM*, system output is characterised by a number of states, which form a system space. The mechanism of *SSM* is most easily introduced by assuming a system space that consists of a finite number of states. Dempster's pioneer work and system view lays the foundation of *SSM*.

Let H_n stand for a singleton state (also termed as hypothesis, assertion or proposition in literature). Without loss of generality, suppose a system space has N singleton states that are mutually exclusive and collectively exhaustive, denoted by $\Theta = \{H_1 \cdots H_n \cdots H_N\}$ with $H_i \cap H_j = \varnothing$ for any $i \neq j$. Probability can be assigned to any singleton states and their subsets. The collection of all singleton states and their subsets in Θ is referred to as the power set of Θ, denoted by 2^{Θ}. Note that the power set of Θ includes empty set $\varnothing$ and system space Θ.

A system output is modelled by a unique set function, called basic probability function, defined as an ordinary probability distribution over the nonempty subsets of Θ as follows.

Definition 1. (*Basic probability function*) A set function $p: 2^{\Theta} \rightarrow [0, 1]$ is called a basic probability function if the following conditions are met:

Table 10
Joint Normalised Likelihood for Both Saliva and Blood.

Likelihood $p_{H_{SBt\theta},(1i,2j)}$	Positive Saliva (x_{11})		Negative Saliva (x_{12})	
	Positive Blood (x_{21})	Negative Blood (x_{22})	Positive Blood (x_{21})	Negative Blood (x_{22})
H_{SBtT}	0.6728	0.2368	0.4337	0.0069
H_{SBtF}	0.0161	0.5263	0.2410	0.6322
H_{SBtU}	0.3111	0.2368	0.3253	0.3608

$$0 \le p(\theta) \le 1, \forall \theta \subseteq \Theta;\ \sum_{\theta \subseteq \Theta} p(\theta) = 1;\ p(\varnothing) = 0 \tag{1}$$

θ can be any singleton state or any subset of singleton states and is referred to as state for short. $p(\theta)$ is the basic probability committed exactly to state θ, which cannot be split into pieces to be assigned to any subsets of θ. In other words, $p(\theta)$ measures the part of evidence that exactly supports state θ as a whole.

The above definition means that one and only one state occurs surely as an outcome when an experiment is conducted. Therefore, the sample space of the probability space, where the experiment is conducted, has $2^N - 1$ states because basic probabilities can be assigned to any state but not to empty set $\varnothing$. For ease of discussion, the above probability space is coined as **S**ystem **O**utput **P**robability **S**pace (*SOPS* for short) in general. If there is a basic probability assigned to any subset of two or more singleton states, or there exists ambiguity, the *SOPS* is referred to as **Dempster's probability space**, to respect Dempster's pioneer work on modelling ambiguity. If basic probabilities are only allowed to be assigned to singleton states, then the *SOPS* is referred to as **Bayes' probability space**, whose sample space has only N singleton states because only a singleton state can be a sure outcome when an experiment is conducted. Bayes' probability space is the ordinary probability space.

A basic probability function constructed in Bayes' probability space can only model randomness among singleton states in the sense that the probability assigned to any singleton state and the probability assigned to its negation always add up to one, leaving no room to model ambiguity among any singleton states. In Dempster's probability space, however, both randomness and ambiguity can be modelled in the same framework because the basic probability assigned to a subset of singleton states as a whole measures the degree to which it is ambiguous or unknown which of the singleton states in the subset may be true.

Given observations of input variables, one or multiple pieces of evidence can be acquired, from which a system output can be generated and in general represented as a basic probability distribution defined as follows.

Definition 2. (*System output*) In *SOPS*, system output y given evidence e as system input is profiled as a basic probability distribution (*bpd* for short) over all states as follows.

$$y(e) = \left\{ (\theta, p(\theta|e)),\ \forall \theta \subseteq \Theta \text{ with } \theta \neq \varnothing \text{ and } \sum_{\theta \subseteq \Theta} p(\theta|e) = 1 \right\} \tag{2}$$

$p(\theta|e)$ is the basic probability, or the conditional probability, that state θ is true given evidence e. Element $(\theta, p(\theta|e))$ represents a judgement or an assertion that θ is true with a probability of $p(\theta|e)$. If $p(\theta|e) > 0$, element $(\theta, p(\theta|e))$ is called a focal element. Since no probability is allowed to be assigned to empty set, there is always $p(\varnothing|e) = 0$. Therefore, $(\varnothing, 0)$ does not need to be included in Equation (2).

In the example of Section 2, let H_1 stand for a singleton state that "a patient has the disease" and H_2 for the other singleton state that "a patient does not have the disease". Then, the system space is defined by $\Theta = \{H_1, H_2\}$. If a test result for a patient is represented by evidence e_1 as system input, a diagnosis result for the patient given the test result is in general represented by the following *bpd* as system output:

$$y(e_1) = \{(H_1, p(H_1|e_1)),\ (H_2, p(H_2|e_1)),\ (\Theta, p(\Theta|e_1))\ \}. \tag{3}$$

where $p(H_1|e_1)$ ($p(H_2|e_1)$) is the basic probability that state H_1 (H_2) is true given evidence e_1 and $p(\Theta|e_1)$ is the basic probability that it is ambiguous or unknown whether the patient has the disease or not given evidence e_1.

System output $y(e_1)$ can be visualised by a probability pie chart, with state H_i represented by the i^{th} sector of the pie chart and $p(H_i|e_1)$ equal to the proportion of the sector area to the total area of the pie chart. If $p(H_1|e_1)$, $p(H_2|e_1)$ and $p(\Theta|e_1)$ are assumed to be 0.6, 0.1, 0.3, respectively, then system output $y(e_1)$ can be visualised by a pie chart as shown in Fig. 1.

Although a *bpd* provides a unique and panoramic description for an output as a whole, uncertainty about each state can be characterised and interpreted in more detail for reasoning and decision making. The system view proposed by Dempster [14] provides

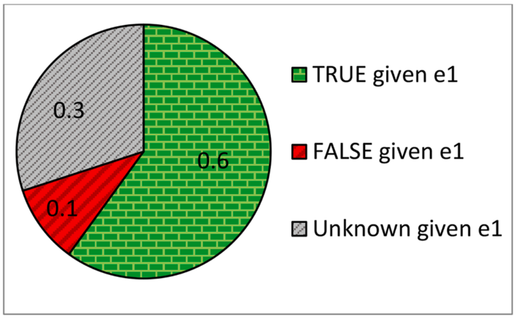

Fig. 1. Pie chart for system output $y(e_1)$.

a basis to describe how uncertainty about a state should be presented and probabilistic operations conducted for inference with both randomness and ambiguity taken into account. In Dempster's system view, uncertainty about state θ is characterised by a triplet $\left(p_\theta^t, p_\theta^f, p_\theta^u\right)$, with p_θ^t measuring the probability that the evidence is for the truth of θ, given by

$$p_\theta^t = \sum_{A\subseteq\theta} p(A|e) \tag{4}$$

That is, p_θ^t accumulates the basic probabilities for all the subsets of states that belong to θ and represents the probability that is unambiguously assigned to θ. As such, p_θ^t is the successor of ordinary probability to which state θ is true.

On the other hand, p_θ^f measures the probability that the evidence is against the truth of θ, or the degree to which θ is believed to be false, calculated by

$$p_\theta^f = p_{\theta^c}^t = \sum_{B\cap\theta=\varnothing} p(B|e) \tag{5}$$

which is the probability for the truth of the negation of θ, denoted by θ^c. p_θ^t and p_θ^f given above are defined so that the assignment of nonzero probability is allowed to be given to "unknown", denoted by p_θ^u and given by

$$p_\theta^u = 1 - p_\theta^t - p_\theta^f \tag{6}$$

In summary, state θ can be interpreted by three probabilities $\left(p_\theta^t, p_\theta^f, p_\theta^u\right)$ assigned to the triad of the state being "true", "false" and "unknown". In other words, probability is not restricted to p_θ^t and p_θ^f with $p_\theta^t + p_\theta^f = 1$ as in Bayes' probability space, but is allowed for $p_\theta^u = 1 - p_\theta^t - p_\theta^f$, to quantify the ambiguity of state θ, so that inference can be conducted with ambiguous information. In *SSM*, while an output is profiled by a unique *bpd*, uncertainty about every state of the *bpd* in the output can be characterised by a $\left(p_\theta^t, p_\theta^f, p_\theta^u\right)$ triplet.

However, while p_θ^t, p_θ^f and p_θ^u are directly interpretable as probabilities and thus useful in decision making, it should be noted that they are not directly usable for probabilistic inference in *SOPS*. This is because the components p_θ^t, p_θ^f and p_θ^u are partial sums of basic probabilities and therefore not *bpd*. As such, the concept of *bpd* is technically fundamental because it is the unique measure of a system output. Only a *bpd* takes the familiar mathematical form of an ordinary probability distribution in *SOPS* and therefore it is only *bpd* or basic probability function that is subject to probabilistic operations and computations in *SOPS* [14].

The above-summarised Dempster's original work and system view on how to handle ambiguity is coined as Dempster's principle that consists of three main components. That is, i> basic probability assigned to state θ cannot be split into pieces to be assigned to any subsets of θ; ii> probabilistic operations and computations can only be conducted on a basic probability function as defined in Equation (2) and implemented in Equation (3); iii> uncertainty about state θ is characterised and interpreted by a triplet $\left(p_\theta^t, p_\theta^f, p_\theta^u\right)$ with Equation (4) to Equation (6) duly satisfied. It is essential that Dempster's principle be strictly followed whenever ambiguity needs to be handled for probabilistic inference.

3.2. Evidence Acquisition Model – EAM

The initial inspiration for creating an Evidence Acquisition Model (*EAM* for short) came from reading the first three chapters of Shafer's book [46], where vivid terms are used to describe that evidence acquired from data or human knowledge should point to different states (hypotheses or propositions) and provide support for or objection against whether a state is true or not. In other words, any acquired evidence should have a relative (not always absolute) role or weight in inference. These descriptions are consistent with the system view that whether a state in a system space is asserted to be true or not depends on whether or not there is evidence pointing towards this state and supporting it. That is, if there is evidence pointing to a state and the evidence is reliable to a certain degree, the state should be true to some degree.

The above system view also means that if a piece of evidence is acquired from a data source and points to a state, it does not always guarantee that the state is necessarily true. As such, evidence should be acquired as system input in association with states, rather than as system output directly profiled by *bpd* on states. This thinking leads to decomposing evidence into evidential elements with each pointing to a state, and to profiling evidence by *bpd* on evidential elements. It is to describe the process of acquiring evidence from a data source and profiling it to form system inputs that an evidence acquisition model (*EAM*) is established in this section.

Let $e_{\theta l}$ be an evidential element pointing exactly to state θ from a data source where input variable x_l takes its values. In other words, $e_{\theta l}$ is the mapping of state θ to the data source. A basic probability function p_l over evidential elements $e_{\theta l}$ $(\theta \subseteq \Theta)$, referred to as *evidence probability function,* is then defined as follows.

Definition 3. (*Evidence probability function*) An evidence probability function p_l for input variable x_l is a basic probability function defined over evidential elements as follows.

$$0 \le p_l(e_{\theta l}) \le 1,\ \forall\theta \subseteq \Theta;\ \sum_{\theta\subseteq\Theta} p_l(e_{\theta l}) = 1;\ p_l(\varnothing) = 0 \tag{7}$$

The above definition requires that one and only one evidential element be taken as the outcome for sure by input variable x_l when an experiment is conducted. The sample space of a probability space where the experiment is conducted has $2^N - 1$ evidential elements because each evidential element points to one and only one non-empty state in *SOPS*. In other words, the new probability space spanned by all the $2^N - 1$ evidential elements is the mapping of *SOPS* to the data source for input variable x_l. For convenience of discussion, this new probability space is coined as **System Input Probability Space** (*SIPS* for short). This means that a *SIPS* should be constructed for each data source where single or multiple input variables take values.

When variable x_l takes a value from its data source, a piece of evidence is acquired that may point to one or multiple states with different probabilities. Let $p_{\theta li} = p_l(e_{\theta l}|x_l = x_{li})$ be the basic probability that state θ is pointed to when input variable x_l takes value x_{li}, or that x_l points to θ at $x_l = x_{li}$ for short. Let e_{li} be a piece of evidence acquired when input variable x_l takes value x_{li}. Then, evidence e_{li} is defined as follows.

Definition 4. (*System input*) A piece of evidence e_{li}, which is taken as system input and acquired when input variable x_l takes value x_{li}, is profiled as a *bpd* over evidential elements by

$$e_{li} = \left\{ (e_{\theta l},\ p_{\theta li} = p_l(e_{\theta l}|x_l = x_{li})),\ \forall \theta \subseteq \Theta \text{ with } \theta \neq \emptyset \text{ and } \sum_{\theta \subseteq \Theta} p_{\theta li} = 1 \right\} \tag{8}$$

Basic probability $p_{\theta li}$ can be calculated by applying the likelihood analysis method [72] as briefly described in Section 2.3. Suppose $c_{\theta li}$ is the likelihood that value x_{li} is observed given evidential element $e_{\theta l}$. $p_{\theta li}$ can then be acquired as normalised likelihood by applying the likelihood method,

$$p_{\theta li} = c_{\theta li} \Big/ \sum_{A \subseteq \Theta} c_{Ali} \qquad \forall \theta \subseteq \Theta \tag{9}$$

For instance, in the example introduced in Section 2, suppose saliva test is denoted by variable x_1 and a positive saliva test result by variable value x_{11}, or $x_1 = x_{11}$. Let e_{11} stand for the evidence that x_{11} is observed, and $e_{H_1 1}$, $e_{H_2 1}$ and $e_{\Theta 1}$ for the evidential elements that saliva test points to the disease being true, false, and unknown, respectively. Given that the basic probabilities for e_{11} were acquired as the normalised likelihoods of Table 8, we have $p_{H_1 11} = 0.627$, $p_{H_2 11} = 0.0535$ and $p_{\Theta 11} = 0.3195$, and can model e_{11} as a *bpd* in *SIPS* as follows:

$$e_{11} = \{(e_{H_1 1},\ 0.627), (e_{H_2 1},\ 0.0535), (e_{\Theta 1},\ 0.3195)\} \tag{10}$$

If variable x_l is discrete, the observation of each of its discrete values can be represented by Definition 4, leading to a discrete *EAM* for probabilistically associating input variable x_l to output variable y, denoted by $E_l = \{e_{l1},\ e_{l2} \cdots, e_{li} \cdots\}$. For example, x_1 has two values: positive saliva test result x_{11} and negative one, denoted by x_{12}. If x_{12} is observed, then a new piece of evidence can be acquired as a *bpd* in the same *SIPS* as for x_1, denoted by e_{12}. From Table 8, basic probabilities for e_{12} are generated by $p_{H_1 12} = 0.0448$, $p_{H_2 12} = 0.6083$ and $p_{\Theta 12} = 0.3469$, respectively, and e_{12} is then given as follows:

$$e_{12} = \{(e_{H_1 1},\ 0.0448), (e_{H_2 1},\ 0.6083), (e_{\Theta 1},\ 0.3469)\} \tag{11}$$

The discrete *EAM* for variable x_1 consists of two piece of evidence and is represented by $E_1 = \{e_{11},\ e_{12}\}$, showing the probabilistic relationships between saliva test and the disease prediction.

Evidence e_{li} can be visualised by a probability pie chart, with $e_{\theta i}$ represented by the i^{th} sector of the pie chart and $p_{\theta li}$ equal to the proportion of the sector area to the total pie chart area. For example, the probability pie charts for e_{11} and e_{12} are shown in Fig. 2.1 and Fig. 2.2, where the green, red and grey sectors of the pie chart represent three evidential elements: $e_{H_1 1}$, $e_{H_2 1}$ and $e_{\Theta 1}$, respectively. For each piece of evidence, the sector areas divided by the total area of the pie chart are equal to the basic probabilities for the three evidential elements.

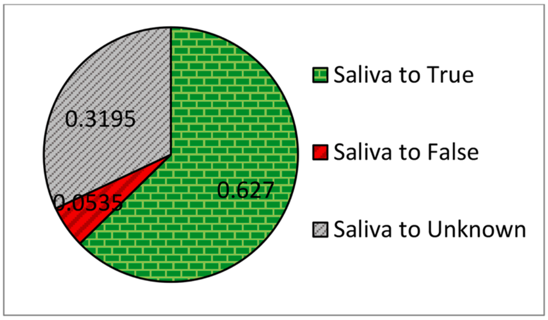

Fig. 2.1. Pie chart for evidence e_{11}.

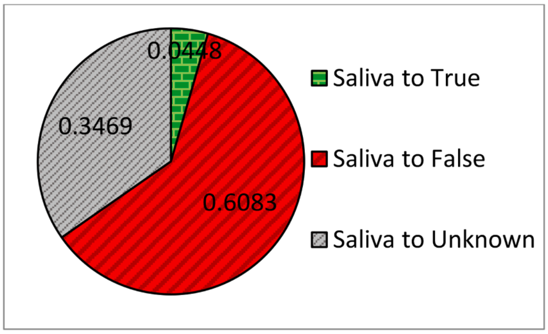

Fig. 2.2. Pie chart for evidence e_{12}.

3.3. Evidential Reasoning Model – ERM

The inspiration for establishing an Evidential Reasoning Model (*ERM* for short) stems from the authors' original work on establishing the *ER* approach for *M*ultiple *A*ttribute *D*ecision *M*aking (*MADM*) [64,66,67]. In *MADM*, the importance of an attribute is measured by its weight, and the assignment of weight depends on the *D*ecision *M*aker's (*DM*'s) preferences; on the other hand, there can be ignorance or ambiguity in assessing alternative actions on attributes due to lack of data or experience. While both the *DM*'s preferences and assessor's ignorance are uncertain in nature, they are different types of uncertainty. The weight of an attribute as perceived by the *DM* and the ignorance or ambiguity incurred in assessing alternative actions on the attribute should therefore be modelled differently.

In the previous section, it was shown how to acquire evidence when an input variable takes a specific value from a data source and how to profile it as a *bpd* over evidential elements that point to different states in *SOPS*. Such acquired evidence provides support for states. If such acquired evidence points to a state to some degree, however, it does not necessarily mean that the state is true to the same degree. This is because no input variable is able to guarantee that every value (judgment or assertion) it takes is always correct without any doubt. As such, an input variable should not always be assumed to have the highest weight, and consequently the evidence acquired when the input variable takes a specific value should not always be regarded as absolutely reliable or trustworthy either. It is to model the weights of input variables and the reliability and trustworthiness of evidence, as well as to enable the combination of multiple pieces of dependent evidence, that *ERM* is established.

As discussed in the previous section, $p_{\theta li} = p_l(e_{\theta l}|x_l = x_{li})$ is the basic probability that variable x_l points to state θ at $x_l = x_{li}$, or that evidence e_{li} points to state θ, which is measured in *SIPS*. However, this does not necessarily mean that θ is true to the same degree of $p_{\theta li}$. What should be considered is the joint event that θ is true while x_l points to θ. This joint event is mathematically represented by the intersection $s_{\theta l} = \theta \cap e_{\theta l}$, with symbol $s_{\theta l}$ standing for the event that variable x_l supports state θ. The probability that x_l supports state θ at $x_l = x_{li}$, or that evidence e_{li} supports θ, is measured by $p(s_{\theta l}|e_{li})$ in *SOPS*. The above discussions lead to the following definition of evidence trustworthiness.

Definition 5. (*Evidence trustworthiness*) The trustworthiness of evidence e_{li} is defined as the total probability that e_{li} supports θ for all $\theta \subseteq \Theta$, denoted by t_{li} as follows:

$$t_{li} = \sum_{\theta \subseteq \Theta} p(s_{\theta l}|e_{li}) \qquad \text{with } 0 \le t_{li} \le 1 \tag{12}$$

The conditional probability that state θ is true given evidence e_{li} is calculated by:

$$p(\theta|e_{li}) = p(s_{\theta l}|e_{li})/t_{li} \tag{13}$$

The system output given evidence e_{li} is then generated as the following *bpd*:

$$y(e_{li}) = \left\{ (\theta, p(\theta|e_{li})),\ \forall \theta \subseteq \Theta \textit{ with } \theta \ne \emptyset \textit{ and } \sum_{\theta \subseteq \Theta} p(\theta|e_{li}) = 1 \right\} \tag{14}$$

Evidence e_{li} is 100% or completely trusted if $t_{li} = 1$, and not trusted at all if $t_{li} = 0$.

The above *bpd* provides a panoramic view of system output $y(e_{li})$ under the two types of uncertainty: randomness and ambiguity, which can be characterised by using a triplet for each state as discussed in Section 3.1. Since t_{li} is the total probability that all states are supported by evidence e_{li}, it measures how much e_{li} is trusted, thus named as the trustworthiness of e_{li}. If $t_{li} < 1$, $(1 - t_{li})$ is the probability that e_{li} is unable to support any state or is untrusted, thus coined as the untrustworthiness of e_{li}. Since $(1 - t_{li})$ actually measures what is left-over by e_{li}'s inability to provide 100% support for all states, it is also referred to as e_{li}'s residual support.

When multiple pieces of evidence acquired from different input variables are combined to infer system output, both the trustworthiness and untrustworthiness of each piece of evidence should be taken into account alongside its randomness and ambiguity. That is, basic probability $p(\theta|e_{li})$ should be discounted by t_{li} and $(1 - t_{li})$ should be considered as well when e_{li} is combined with other evidence. Basic probability $p(\theta|e_{li})$ discounted by t_{li} measures evidence e_{li}'s support for state θ, and is denoted by:

$$\widetilde{m}_{\theta li} = t_{li}p(\theta|e_{li}) = p(s_{\theta l}|e_{li}) \qquad \forall\theta \subseteq \Theta \tag{15}$$

It has been argued from different perspectives such as *MADM* [67,71] that it is irrational and violates the likelihood principle [5] to assign the residual support or untrustworthiness to any state of a system space. On the other hand, it is rational and necessary to retain it to the powerset as untrustworthiness is a new dimension of uncertainty, different from randomness or ambiguity, and can only be robustly modelled in an augmented *SOPS* by taking the powerset as an augmented state to hold the new type of uncertainty.

For ease of discussion, the augmented *SOPS* is referred to as the ***ER* Probability Space** (*ERPS* for short). In essence, a piece of evidence is profiled by two main parts in *ERPS*. In the first part, the basic probabilities of the evidence are each discounted by the same factor (its trustworthiness) and assigned to their respective states in *SOPS*, thus with their probability meanings kept intact; in the second part, the untrustworthiness of the evidence is not assigned to any state but is retained to the powerset as a whole and kept intact, ready for combination with other evidence. Such retainment is appropriate, rigorous, logical and convenient for evidence combination.

- First, the powerset is built upon the same system space in that it consists of only and all the states of the system space with no alien state fabricated. It is therefore appropriate to hold the untrustworthiness of evidence to the powerset as a whole.
- Second, the retainment is rigorous since a piece of evidence profiled by Equation (15) with untrustworthiness assigned to the powerset as a whole holds the same probability meanings (or relative frequency) as the original *bpd* of the evidence measured in *SOPS* because

$$\widetilde{m}_{Ali}/\widetilde{m}_{Bli} = p(A|e_{li})/p(B|e_{li}) \qquad \forall A,\ B \subseteq \Theta \tag{16}$$

- Third, it is logical to retain the untrustworthiness to the powerset as a whole because this gives due respect to the meaning of the untrustworthiness of evidence that is incurred due to the inability of the evidence to provide 100% support for each of all states. As such, the untrustworthiness should not be assigned to any specific state with others ignored. On the other hand, the untrustworthiness of evidence in nature measures the inability of the evidence to decide on its own which state the untrustworthiness should be assigned to. Retaining the untrustworthiness to the powerset makes it allocatable to any state during the process of combining the evidence with other evidence, because the conjunction of the powerset with any of its constituent states is the constituent state.
- Finally, the retainment enables untrustworthiness to be constructed entirely on the basis of the evidence, independent of any other evidence. Such independence is useful as this makes it convenient to combine the evidence with other evidence in the process of probabilistic operations such as calculating joint probabilities.

The above discussions lead to the construction of the following *ER* probability function.

Definition 6. (*ER probability function*) An *ER* probability function *P* is an augmented basic probability function constructed over all the states and the powerset of the system space by

$$0 \le P(\theta) \le 1,\ \forall\theta \subseteq \Theta;\ \sum_{\theta\subseteq\Theta} P(\theta) \le 1;\ P(2^{\Theta}) = 1 - \sum_{\theta\subseteq\Theta} P(\theta);\ P(\varnothing) = 0 \tag{17}$$

$P(\theta)$ is the basic probability that state θ is supported and $P(2^{\Theta})$ is the remaining probability leftover after all the states are supported, which cannot be broken down into pieces to be assigned to any of the states but is kept intact and retained to the powerset as a whole.

The above definition ensures that $0 \le P(2^{\Theta}) \le 1$ and $\sum_{\theta\subseteq\Theta} P(\theta) + P(2^{\Theta}) = 1$ and allows evidence acquired from any input variable to be profiled over all the states and the powerset of the system space, so that evidence from all input variables is profiled and compared on the same basis, whatever values the input variables take. The pieces of evidence acquired when input variables take any values and consistently measured by Definition 6 can then be combined to infer system outputs.

The basic probability $P(\theta|e_{li})$ that state θ is supported at $x_l = x_{li}$, or evidence e_{li} supports state θ, is given by a joint probability, or $P(\theta|e_{li}) = p(s_{\theta l}|e_{li})$ with *p* being a basic probability function defined in *SOPS*. Since e_{li} is acquired in *SIPS* and cannot be directly used to calculate $p(s_{\theta l}|e_{li})$, it needs to be projected to *SOPS* in a procedure discussed as follows. First, the joint probability is rewritten as follows:

$$P(\theta|e_{li}) = p(s_{\theta l}|e_{li}) = p(\theta \cap e_{\theta l}|e_{li}) = p(\theta|e_{\theta l})p(e_{\theta l}|e_{li}) \tag{18}$$

In the above equation, $p(\theta|e_{\theta l})$ is the conditional probability that state θ is true given that input variable x_l points to θ, defined as the weight of x_l in support for θ as follows.

Definition 7. (*Variable weight*) The weight of input variable x_l in support for state θ is defined as the conditional probability that state θ is true given that x_l points to θ, given by

$$w_{\theta l} = p(\theta|e_{\theta l}) \tag{19}$$

with $0 \leq w_{\theta l} \leq 1$. $w_{\theta l} = 1$ means that x_l is the most important when pointing to θ, and $w_{\theta l} = 0$ means that x_l is the least important when pointing to θ.

Weight $w_{\theta l}$ measures the degree to which state θ is true if x_l points to θ, without referring to any specific value that x_l may take. In another word, $w_{\theta l}$ describes the ability of x_l to provide correct judgement or assertion about the truth of θ, Therefore, $w_{\theta l}$ should in principle be estimated by analysing the ability of x_l to provide correct judgement or assertion for θ, whatever specific values x_l may take. Under this principle, an optimisation model will be developed in Section 5.1 to estimate $w_{\theta l}$ from observation data.

In the previous section, x_l is characterized by evidence probability function p_l, with $p_{\theta li} = p_l(e_{\theta l}|x_l = x_{li})$ being the basic probability that input variable x_l points to θ at $x_l = x_{li}$, which is acquired in the *SIPS* from the data source where x_l takes its values. On the other hand, $p(e_{\theta l}|e_{li})$ is the basic probability that state θ is true given evidence e_{li}, with p being the basic probability function constructed in *SOPS*. Therefore, $p_{\theta li}$ needs to be projected to $p(e_{\theta l}|e_{li})$ for all $\theta \subseteq \Theta$. In other words, evidence e_{li} acquired in *SIPS* needs to be projected to *SOPS*. This projection must follow the likelihood principle [5] in that after the projection $p(e_{\theta l}|e_{li})$ should hold the same evidential meanings as what $p_{\theta li}$ holds for all $\theta \subseteq \Theta$. That is, $p(e_{\theta l}|e_{li})$ should be proportional to $p_{\theta li}$ for all $\theta \subseteq \Theta$ as follows:

$$p(e_{\theta l}|e_{li}) = \omega_{li} p_l(e_{\theta l}|x_l = x_{li}) = \omega_{li} p_{\theta li} \forall \theta \subseteq \Theta \tag{20}$$

ω_{li} is the proportion that evidence e_{li} is projected to *SOPS*, referred to as projection rate.

Projection rate ω_{li} is a positive constant that does not change with $\theta \subseteq \Theta$, and is determined uniquely by the reliability of evidence e_{li} and the weights of input variable x_l as follows, so that Equation (17) of Definition 6 is satisfied:

$$\omega_{li} = 1 \Bigg/ \left(\sum_{\theta \subseteq \Theta} p_{\theta li} w_{\theta l} + 1 - r_{li} \right) \tag{21}$$

In the above equation for determining the projection rate, $\sum_{\theta \subseteq \Theta} p_{\theta li} w_{\theta l}$ is the expected weight of evidence e_{li}, and r_{li} is the reliability of e_{li} defined as follows.

Definition 8. (*Evidence reliability*) Parameter r_{li} given in Equation (21) measures the ability of evidence e_{li} to provide correct outcomes or how reliable e_{li} is in its support for states, referred to as the reliability of e_{li} with $0 \leq r_{li} \leq 1$. Evidence e_{li} is the most reliable or fully reliable if r_{li}=1 and the least reliable or fully unreliable if r_{li}=0, with the former meaning that whatever e_{li} stands for is always correct and the latter meaning exactly the opposite.

We can now calculate the basic probability that e_{li} supports θ as follows:

$$\begin{aligned} \widetilde{m}_{\theta li} &= P(\theta|e_{li}) = p(s_{\theta l}|e_{li}) = p(\theta|e_{\theta l}) p(e_{\theta l}|e_{li}) \\ &= p(\theta|e_{\theta l}) \omega_{li} p_l(e_{\theta l}|x_l = x_{li}) = \omega_{li} m_{\theta li} \qquad \forall \theta \subseteq \Theta \end{aligned} \tag{22}$$

$$m_{\theta li} = p(\theta|e_{\theta l}) p_l(e_{\theta l}|x_l = x_{li}) = w_{\theta l} p_{\theta li} \qquad \forall \theta \subseteq \Theta \tag{23}$$

$m_{\theta li}$ is the weighted likelihood that e_{li} points to θ, referred to as probability mass. Note that there is always $0 \leq m_{\theta li} \leq 1$ because $0 \leq w_{\theta l} \leq 1$ and $0 \leq p_{\theta li} \leq 1$ for any $\theta \subseteq \Theta$. Also, there must be $0 \leq \widetilde{m}_{\theta li} = \omega_{li} m_{\theta li} \leq 1$ for any $\theta \subseteq \Theta$ because $p(s_{\theta l}|e_{li})$ is probability, which means that there must not be $\sum_{\theta \subseteq \Theta} w_{\theta l} p_{\theta li} = 0$ when r_{li}=1. In other words, if a piece of evidence is fully reliable, its expected weight must not be zero. In general, reliability can be different from expected weight for any evidence. If the former is larger than (less than or equal to) the latter, there is $\omega_{li} > (< or =)\ 1$ and the support of evidence e_{li} for each state is then amplified (shrunk or kept unchanged) from its probability mass for the state.

The trustworthiness and untrustworthiness of evidence e_{li} can then be rewritten as follows.

$$t_{li} = \sum_{\theta \subseteq \Theta} p(s_{\theta l}|e_{li}) = \sum_{\theta \subseteq \Theta} \omega_{li} m_{\theta li} = \omega_{li} \sum_{\theta \subseteq \Theta} p_{\theta li} w_{\theta l} = \frac{\sum_{\theta \subseteq \Theta} p_{\theta li} w_{\theta l}}{\sum_{\theta \subseteq \Theta} p_{\theta li} w_{\theta l} + 1 - r_{li}} \tag{24}$$

$$1 - t_{li} = 1 - \frac{\sum_{\theta \subseteq \Theta} p_{\theta li} w_{\theta l}}{\sum_{\theta \subseteq \Theta} p_{\theta li} w_{\theta l} + 1 - r_{li}} = \frac{1 - r_{li}}{\sum_{\theta \subseteq \Theta} p_{\theta li} w_{\theta l} + 1 - r_{li}} = \omega_{li}(1 - r_{li}) \tag{25}$$

$$\widetilde{m}_{2^{\theta} li} = P(2^{\Theta}|e_{li}) = 1 - \sum_{\theta \subseteq \Theta} p(s_{\theta l}|e_{li}) = 1 - t_{li} = \omega_{li}(1 - r_{li}) \tag{26}$$

Note from the above equations that the trustworthiness of evidence e_{li} is positively related to its expected weight, and the untrustworthiness of e_{li} is positively related to its unreliability. Equation (25) shows that there is $(1 - t_{li}) = 0$ when $r_{li} = 1$. That is, evidence e_{li} is 100% or completely trusted if it is fully reliable. In general, if $r_{li} > (< or =)\ \sum_{\theta \subseteq \Theta} p_{\theta li} w_{\theta l}$, there is $\omega_{li} > (< or =)\ 1$ and the untrustworthiness of evidence e_{li} is amplified from (shrunk from or is the same as) its unreliability $(1 - r_{li})$.

Also note that if $w_{\theta l} = w_l$ for all $\theta \subseteq \Theta$, we will have $\sum_{\theta \subseteq \Theta} p_{\theta li} w_{\theta l} = w_l$, thus resulting in $t_{li} = \frac{w_l}{w_l + 1 - r_{li}}$ and $1 - t_{li} = \frac{1 - r_{li}}{w_l + 1 - r_{li}}$, with $t_{li} = 1$ if $r_{li} = 1$. This leads to another interpretation of t_{li} as the hybrid weight of e_{li} that takes into account a mix of weight and reliability and is denoted by $\widetilde{w}_{li}$ [71].

Given the above discussions, we can now construct a *bpd* for evidence e_{li} in *ERPS*, referred to as an *ER* probability distribution as follows.

Definition 9. (*ER probability distribution*) An *ER* probability distribution for evidence e_{li} is an augmented *bpd* defined over all the states and the powerset of the system space, denoted by m_{li} or $ER(e_{li})$, to profile e_{li}'s support for all states as well as the powerset as follows:

$$\begin{aligned} m_{li} = ER(e_{li}) &= \{(\theta,\ P(\theta|e_{li})),\ \forall\theta\subseteq\Theta \text{ with } \theta\neq\varnothing;\ (2^{\Theta},\ P(2^{\Theta}|e_{li}))\} \\ &= \{(\theta,\ \omega_{li}m_{\theta li}),\ \forall\theta\subseteq\Theta \text{ with } \theta\neq\varnothing;\ (2^{\Theta},\ 1-t_{li})\} \end{aligned} \tag{27}$$

In Definition 9, the augmented element $(2^{\Theta},\ 1-t_{li})$ represents the residual support of evidence e_{li} leftover after e_{li} supports all states, and is retained to the powerset, coined as the support for the powerset for short, with $(1-t_{li})$ being the probability of the residual support or that e_{li} is untrusted for any specific states. As such, $(1-t_{li})$ cannot be broken down into pieces to be redistributed to any specific state but is retained to the powerset and kept intact, ready for combination with other evidence. Definition 9, Equation (24) and Equation (25) show that reliability r_{li} describes the ability of evidence e_{li} to ensure the correctness or trustworthiness of an outcome (conclusion or assessment) in the sense that e_{li} is completely trusted if it is fully reliable and e_{li}'s untrustworthiness is positively related to its unreliability.

Definition 9 provides a unified framework, where the untrustworthiness of e_{li} alongside its randomness and ambiguity are all modelled by a single basic probability function defined on the same set of states and the powerset, thus enabling multiple pieces of evidence to be consistently modelled and compared for subsequent combination.

If input variable x_l is discrete, a piece of evidence is acquired when each of its discrete values is observed, which can be projected to *SOPS* and modelled by Equation (27). This will lead to an *ERM* for x_l, denoted by $ERM_l = \{m_{l1}, m_{l2}, \cdots, m_{li}, \cdots\}$, representing the probabilistic relationships between input variable x_l and output y in *ERPS*.

In the example of Section 2, if the weight of the first input variable x_1 (saliva test) is assumed to be 0.9 whether it points to the disease being true, false or unknown, and the reliability of evidence e_{11} (positive saliva test result) is also assumed to be 0.9, that is $w_{H_1 1} = w_{H_2 1} = w_{\Theta 1} = 0.9$ and $r_{11} = 0.9$, leading to $\omega_{11} = 1$, then from Equation (22), Equation (23) and Equation (26) we get the following joint probabilities for e_{11}.

$$\widetilde{m}_{H_1 11} = \omega_{11} w_{H_1 1} p_{H_1 11} = 1 \times 0.9 \times 0.627 = 0.5643$$

$$\widetilde{m}_{H_2 11} = \omega_{11} w_{H_2 1} p_{H_2 11} = 1 \times 0.9 \times 0.0535 = 0.0481$$

$$\widetilde{m}_{\Theta 11} = \omega_{11} w_{\Theta 1} p_{\Theta 11} = 1 \times 0.9 \times 0.3195 = 0.2876$$

$$\widetilde{m}_{2^{\theta} 11} = \omega_{11}(1 - r_{11}) = 1 \times (1 - 0.9) = 0.1$$

The *ER* probability distribution for evidence e_{11} is then modelled by m_{11} as follows:

$$\begin{aligned} m_{11} &= \left\{(H_1,\ \widetilde{m}_{H_1 11}),\ (H_2,\ \widetilde{m}_{H_2 11}), (\Theta,\ \widetilde{m}_{\Theta 11});\ \left(2^{\theta},\ \widetilde{m}_{2^{\theta} 11}\right)\right\} \\ &= \left\{(H_1,\ 0.5643),\ (H_2,\ 0.0481), (\Theta,\ 0.2876);\ \left(2^{\theta},\ 0.1\right)\right\} \end{aligned} \tag{28}$$

Similarly, from Equation (11), the *ER* probability distribution for evidence e_{12}, with $r_{12} = 0.9$ and thus $\omega_{12} = 1$, is modelled by m_{12} as follows:

$$\begin{aligned} m_{12} &= \left\{(H_1,\ \widetilde{m}_{H_1 12}),\ (H_2,\ \widetilde{m}_{H_2 12}), (\Theta,\ \widetilde{m}_{\Theta 12});\ \left(2^{\theta},\ \widetilde{m}_{2^{\theta} 12}\right)\right\} \\ &= \left\{(H_1,\ 0.0403),\ (H_2,\ 0.5475), (\Theta,\ 0.3122);\ \left(2^{\theta},\ 0.1\right)\right\} \end{aligned} \tag{29}$$

The *ERM* for saliva test (x_1) is given by $ERM_1 = \{m_{11}, m_{12}\}$, measuring the probabilistic relationships between saliva test and the disease diagnosis in *ERPS*.

The *ER* probability distribution (*ER-pd*) of evidence e_{li} can be visualised by a probability pie chart, where state θ is a sector of the pie chart, and $\widetilde{m}_{\theta li}$ is equal to the proportion of the sector area to the total area of the pie chart. A special circle around the edge of the pie

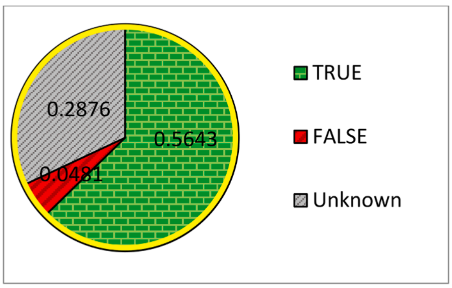

Fig. 3.1. *ER-pd* for evidence e_{11}.

chart is used to represent the residual support or untrustworthiness of e_{li}, with its size determined by $\widetilde{m}_{2^\theta li}$ that is equal to the area of the circle divided by the total area of the pie chart. The pie charts for the *ER* probability distributions of evidence e_{11} and evidence e_{12} are shown in Figs. 3.1 and 3.2, respectively. The circle for each piece of evidence is drawn in yellow, and this part cannot be distributed to any state or any sector of the pie chart by the evidence alone.

An *ERM* can be constructed for each input variable in a similar way. For example, blood test (x_2) is another input variable. If m_{21} and m_{22} are the *ER* probability distributions for the two pieces of evidence, acquired from the observation of a positive or negative blood test result, an *ERM* for x_2 can be constructed and represented by $ERM_2 = \{m_{21}, m_{22}\}$ in the same way as for x_1. If $w_{H_1 2} = w_{H_2 2} = w_{\Theta 2} = 0.95$ and $r_{21} = r_{22} = 0.95$, leading to $\omega_{21} = \omega_{22} = 1$, then *ER* probability distributions for evidence e_{21} and evidence e_{22} are modelled by m_{21} and m_{22}:

$$m_{21} = \left\{(H_1,\ \widetilde{m}_{H_1 21}),\ (H_2,\ \widetilde{m}_{H_2 21}), (\Theta,\ \widetilde{m}_{\Theta 21});\ \left(2^\theta,\ \widetilde{m}_{2^\theta 21}\right)\right\} \\ = \left\{(H_1,\ 0.599),\ (H_2,\ 0.0317), (\Theta,\ 0.3193);\ \left(2^\theta,\ 0.05\right)\right\} \tag{30}$$

$$m_{22} = \left\{(H_1,\ \widetilde{m}_{H_1 22}),\ (H_2,\ \widetilde{m}_{H_2 22}), (\Theta,\ \widetilde{m}_{\Theta 22});\ \left(2^\theta,\ \widetilde{m}_{2^\theta 22}\right)\right\} \\ = \left\{(H_1,\ 0.039),\ (H_2,\ 0.5969), (\Theta,\ 0.3141);\ \left(2^\theta,\ 0.05\right)\right\} \tag{31}$$

3.4. *Illustration of the MAKER Framework*

The three models (*SSM, EAM* and *ERM*) discussed in Sections 3.1, 3.2 and 3.3 constitute the unique *MAKER* framework for characterising probabilistic relationships between input and output variables to enable augmented probabilistic inference. The following example shows how the three models can be constructed and applied in a decision situation under uncertainty.

Example 2. The jury trial is a longstanding part of the criminal justice system in England and Wales, among other countries. The jury consists of 9 to 12 members of the public, and a minimum of 9 jurors are required to deliver its verdict. While criminal trials require a jury to be satisfied with a guilty verdict beyond a reasonable doubt, in civil trials the jury must be satisfied on the balance of probabilities, that is, the defendant's guilt is more probable than not.

In this example, we consider three hypothetical trial cases. In the first case, the jury with 9 jurors reaches the guilty verdict unanimously. In the second case, the jury with 12 jurors considers its verdict with 10 of the 12 jurors pointing to guilty and the other 2 jurors to not guilty. The third case is more complicated, with 9 of the 12 jurors pointing to guilty, the 10$^{\text{th}}$ juror only 50% pointing to guilty with the other 50% being unknown, and the 11$^{\text{th}}$ and 12$^{\text{th}}$ juror each 50% pointing to not guilty with the other 50% being unknown. Suppose every juror is given a weight of 50% as a measure of their perceived ability to make correct judicial judgement and every juror's assessment or conclusion in each of the three cases is regarded 50% reliable. Suppose every juror provides her judgment independent of other jurors in the sense that the juror's judgment does not change whether other jurors' judgments are known to the juror or not. What are the probabilities of the jury's verdict in each case?

The process of reaching a verdict in each of the three cases is regarded as a decision system, with each having two singleton system states: guilty or not guilty, represented by H_1 and H_2, so each system space is defined by $\Theta = \{H_1, H_2\}$. The verdict of the jury is the system output, and each juror's conclusion is a system input, so the system has one output for each of the three trial cases and nine inputs for case 1 and twelve inputs for cases 2 and 3.

The three cases are analysed separately to show how to construct *SSM, EAM* and *ERM* and how to apply the *ER* rule to analyse these cases, which is applicable in each case as the inputs are assumed to be independent of each other.

The analysis of the first trial case is summarised as follows. Since each of the 9 jurors points to H_1 only, the conclusion of each juror can be profiled as an ordinary probability distribution, so the *SOPS* of this case is Bayes' probability space. An output can be represented as the following ordinary *bpd* in general:

$$y(e) = \{(H_1, p(H_1|e)),\ (H_2, p(H_2|e))\} \tag{32}$$

where e is the combined evidence representing the verdict of the jury.

Let $e_{H_1 l}$ ($e_{H_2 l}$) be the evidential element that the l^{th} juror points to a guilty (not guilty) verdict. The l^{th} juror's guilty conclusion is

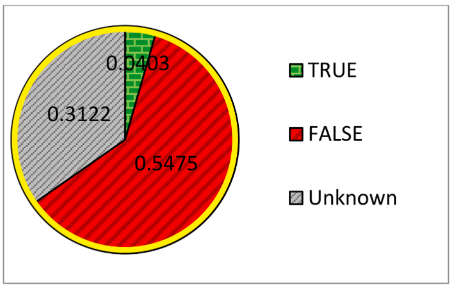

Fig. 3.2. *ER-pd* for evidence e_{12}.

acquired as a piece of evidence e_l and profiled as the following *bpd* over the evidential elements:

$$e_l = \{(e_{H_1 l}, p_l(e_{H_1 l}) = 1)\} \qquad l = 1, \cdots, 9 \tag{33}$$

where p_l is the evidence probability function constructed in the l^{th} juror's *SIPS*.

Let r_l be the reliability of the l^{th} juror's conclusion and $w_{H_1 l}$ ($w_{H_2 l}$) the weights of the l^{th} juror for guilty (not guilty) verdict for each case, respectively. We have $w_{H_n l} = 0.5$ for $n = 1, 2$ and $r_l = 0.5$, resulting in $\omega_l = 1$ for all $l = 1, \cdots, 9$ from Equation (21). The l^{th} juror's conclusion is then projected from *SIPS* to *SOPS* as the following *bpd*:

$$y(e_l) = \{(H_1, p(H_1|e_l) = 1)\} \text{ with } w_{H_1 l} = 0.5 \text{ and } r_l = 0.5 \text{ for } l = 1, \cdots, 9 \tag{34}$$

Note that $p(H_1|e_l) = \omega_l p_l(e_{H_1 l})$ from Equation (20).

To reach the jury's verdict, the l^{th} juror's conclusion is modelled as the following augmented *bpd* in *ERPS* for all $l = 1, \cdots, 9$ from Equation (22) and Equation (26), given $\omega_l = 1$ and $1 - t_l = \omega_l(1 - r_l) = 0.5$ for all l=1, ⋯, 9:

$$m_l = ER(e_l) = \{(H_1, \tilde{m}_{H_1 l} = m_{H_1 l} = 0.5); (2^\Theta, \ \tilde{m}_{2^\Theta l} = m_{2^\Theta l} = 1 - t_l = 0.5)\} \tag{35}$$

The *ER* rule is then applied to combine the 9 jurors' conclusions to reach the jury's verdict. For example, the first two jurors' conclusions are combined by applying the *ER* rule in *ERPS* as follows:

$$ER(e_1 \wedge e_2) = \{(H_1, m_{H_1 12} = 0.75); (2^\Theta, \ m_{2^\Theta 12} = 0.25)\} \tag{36}$$

$ER(e_1 \wedge e_2)$ is then combined with $ER(e_3)$ to generate $ER(e_1 \wedge e_2 \wedge e_3)$. This recursive process is repeated until all 9 jurors' conclusions are combined, leading to the following augmented *bpd*, with $e_{1\overline{\wedge} l} = (e_1 \wedge \cdots \wedge e_l)$:

$$ER(e_{1\overline{\wedge}9}) = \{(H_1, m_{H_1(1\overline{\wedge}9)} = 0.998); (2^\Theta, \ m_{2^\Theta(1\overline{\wedge}9)} = 0.002)\} \tag{37}$$

Use Equation (A.2.9) in [71] to calculate the joint conditional probabilities of the jury's verdict and profile it as the following *bpd* in *SOPS*.

$$y(e_{1\overline{\wedge}9}) = \left\{ \left(H_1, p_{H_1(1\overline{\wedge}9)} = \frac{m_{H_1(1\overline{\wedge}9)}}{1 - m_{2^\Theta(1\overline{\wedge}9)}} = \frac{0.998}{1 - 0.002} = 1 \right) \right\}$$

with the trustworthiness of the *bpd* calculated by

$$t_{1\overline{\wedge}9} = 1 - m_{2^\Theta(1\overline{\wedge}9)} = 0.998 \tag{38}$$

The above analysis results for the first trial case assert that the jury's verdict is 100% guilty with a trustworthiness of 99.8%. In other words, if each of the 9 juror's guilty conclusions is assumed to be 50% reliable and all of them are independent of each other, the jury's guilty verdict can be 99.8% trusted, or there is a 99.8% chance that the jury's guilty verdict is correct.

In the second trail case, there are 12 jurors with 10 of them pointing to guilty but the other 2 jurors to not guilty. Each of the 10 jurors' guilty conclusions is profiled as an augmented *bpd* in *ERPS* in the same way as in Equation (35), but the other 2 juror's not guilty conclusions are each profiled as the following augmented *bpd* in *ERPS*:

$$ER(e_l) = \{(H_2, 0.5); (2^\Theta, \ 0.5)\} \qquad \text{for } l = 11, \ 12 \tag{39}$$

Applying the *ER* rule to combine the 10 jurors' augmented *bpd* $ER(e_l)$ for $l = 1, \cdots, 10$ with each of them identical to Equation (35), followed by $ER(e_{11})$ and $ER(e_{12})$ with each given in Equation (39), leads to the following augmented *bpd* in *ERPS:*

$$ER(e_{1\overline{\wedge}12}) = \{(H_1, 0.9961), (H_2, 0.0029); (2^\Theta, \ 0.001)\} \tag{40}$$

The above augmented *bpd* is then used to generate the conditional probabilities of the jury's verdict profiled as the following *bpd* in *SOPS:*

$$y(e_{1\overline{\wedge}12}) = \{((H_1, 0.9971), (H_2, 0.0029))\} \text{ with } t_{1\overline{\wedge}12} = 0.999 \tag{41}$$

The above analysis results for the second trial case assert that the jury's verdict is 99.71% guilty and 0.29% not guilty with a trustworthiness of 99.9%.

The third trial case is more complicated and different from the second one in that there are three jurors who each have unknown in their conclusions. Due to this, the *SOPS* of this case is Dempster's probability space; an output is in general represented as the following *bpd*:

$$y(e) = \{(H_1, p(H_1|e)), \ (H_2, p(H_2|e)), (\Theta, p(\Theta|e))\} \tag{42}$$

The first 9 jurors' conclusions are each profiled as augmented *bpd* with the focal elements identical to Equation (35). The 10^{th} juror's conclusion is profiled as follows:

$$ER(e_{10}) = \{(H_1, 0.25), (\Theta, 0.25); (2^\Theta, \ 0.5)\} \tag{43}$$

The 11^{th} and 12^{th} juror's conclusions are profiled as the following augmented *bpd:*

$$ER(e_l) = \{(H_2, 0.25), (\Theta, 0.25); (2^{\Theta},\ 0.5)\} \text{ for } l = 11,\ 12 \tag{44}$$

Applying the *ER* rule to combine the first 9 jurors' augmented *bpd* $ER(e_l)$ for $l = 1, \cdots 9$ with each identical to Equation (35) and then the 10th juror's augmented *bpd* $ER(e_{10})$ of Equation (43), followed by $ER(e_{11})$ of Equation (44) and finally $ER(e_{12})$ of Equation (44), leads to the following augmented *bpd:*

$$ER(e_{1\overline{\wedge}12}) = \{(H_1, 0.9974), (H_2, 0.0012), (\Theta, 0.001); (2^{\Theta},\ 0.0004)\} \tag{45}$$

The above augmented *bpd* is then used to generate the conditional probabilities of the jury's verdict profiled as the following *bpd* in *SOPS:*

$$y(e_{1\overline{\wedge}12}) = \{(H_1, 0.9978), (H_2, 0.0012), (\Theta, 0.001)\} \text{ with } t_{1\overline{\wedge}12} = 0.9996 \tag{46}$$

The above analysis results for the third trial case assert that the jury's verdict is 99.78% guilty, 0.12% not guilty and 0.1% unknown with a trustworthiness of 99.96%.

If the jury can reach the guilty verdict in the second case, the jury should also consider reaching the guilty verdict in the third case. This is because the jury's verdict has a higher probability of guilty and a lower probability of not guilty with a higher trustworthiness in the third case than in the second case.

In practice, the jury is allowed to reach a guilty verdict in cases 1 and 2 [40]. Although in these two cases each juror's conclusion is certain and the type of *SOPS* is Bayes' probability space, it is still necessary to combine the jurors' conclusions in *ERPS* to generate the jury's verdict because of the limited weight assigned to each juror and the inherent unreliability of each juror's conclusion. In case 3, some jurors' ambiguous judgments make it necessary to measure the jury's verdict in Dempster's probability space. There is no guideline in practice on whether the jury is allowed to reach the guilty verdict in case 3. However, the above analyses show that the jury should be satisfied with the probabilities of the guilty verdict in case 3 if they are satisfied with the probabilities of the guilty verdict in case 2.

4. Conjunctive *MAKER* rule and special cases

In the last sections, the *ER* rule was applied to combine multiple pieces of independent evidence. Evidence acquired from data, however, is not independent of but dependent on each other in general. The question is how to combine dependent evidence to enable probabilistic inference from system inputs to output. In this section, we attempt to address this question.

4.1. Interrelationships among input variables

When multiple input variables are not independent of each other, their interrelations, as well as evidence acquired from them, need to be described explicitly. If there is a common data source where input variables jointly take values, their interrelationships can be captured by calculating their joint probabilities, as discussed below.

First, the following symbols originally defined in the previous sections are now restated to facilitate the descriptions of interrelationships between input variables x_l and x_m. For x_l, e_{li} is the evidence acquired at $x_l = x_{li}$, e_{Al} the evidential element that points to state A from the l^{th} *SIPS* of a data source where x_l takes values, p_l the evidence probability function defined in the l^{th} *SIPS*, and $p_{Ali} = p_l(e_{Al}|x_l = x_{li})$ the probability that state A is pointed to given $x_l = x_{li}$ or that e_{li} points to state A. For x_m, e_{mj} is the evidence acquired at $x_m = x_{mj}$, e_{Bm} the evidential element that points to state B from the m^{th} *SIPS* of a data source where x_m takes values, p_m the evidence probability function defined in the m^{th} *SIPS*, and $p_{Bmj} = p_m(e_{Bm}|x_m = x_{mj})$ the probability that state B is pointed to given $x_m = x_{mj}$ or that e_{mj} points to state B.

To describe the interrelationship between x_l and x_m and between e_{li} and e_{mj}, let $e_{\theta lm}$ be the evidential element that points to state θ from the *SIPS* of a common data source where both variables x_l and x_m take their values, p_{lm} the evidence probability function constructed in the common *SIPS*, and

$$p_{Ali,Bmj} = p_{lm}\left((e_{Alm}|x_l = x_{li}) \cap \left(e_{Blm}|x_m = x_{mj}\right)\right) \forall A, B \subseteq \Theta \tag{47}$$

the probability that state A is pointed to given $x_l = x_{li}$ and state B is pointed to given $x_m = x_{mj}$. If $c_{Ali,Bmj}$ is the likelihood that x_{li} is observed given e_{Alm} and x_{mj} is observed given e_{Blm} with $A \cap B = \theta$, the normalised likelihood that state A is pointed to given $x_l = x_{li}$ and state B is pointed to given $x_m = x_{mj}$ can be calculated using the likelihood analysis method [72] as follows:

$$p_{Ali,Bmj} = c_{Ali,Bmj} \Big/ \sum_{C \cap D = \theta \subseteq \Theta} c_{Cli,Dmj} \qquad \forall A, B \subseteq \Theta \tag{48}$$

For instance, in the example of Section 2, the probability that a combination of positive saliva test result ($x_1 = x_{11}$) and positive blood test result ($x_2 = x_{21}$) points to the disease being true is given by the joint normalised likelihood $p_{H_1 11, H_1 21} = p_{H_1(11,21)} = 0.6728$ from Table 10.

Given the above discussions, we are now able to measure the interrelationships between two pieces of evidence by defining the following interdependence index.

Definition 10. (*Interdependence index*) The interrelationship between evidence e_{li} and e_{mj} for any $A, B \subseteq \Theta$ with $A \cap B = \theta$ is described by the following interdependence index:

$$\overline{\alpha}_{Ali,Bmj} = \begin{cases} 0 \text{ if } p_{Ali} = 0 \text{ or } p_{Bmj} = 0 \\ p_{Ali,Bmj} \Big/ \left(p_{Ali} p_{Bmj}\right) \text{ otherwise} \end{cases} \tag{49}$$

The interdependence index is non-negative and describes how two pieces of evidence are interrelated to each other. If $p_{Ali,Bmj}$=0, we have $\overline{\alpha}_{Ali,Bmj} = 0$. If this is the case for any $A,\ B \subseteq \Theta$ with $A \cap B = \theta$, it is said that e_{li} and e_{mj} are evidentially exclusive of each other (see also Definition 13). If $p_{Ali,Bmj} = p_{Ali}p_{Bmj}$, we have $\overline{\alpha}_{Ali,Bmj} = 1$. If this is the case for any $A,\ B \subseteq \Theta$ with $A \cap B = \theta$, it is said that e_{li} and e_{mj} are evidentially independent of each other.

If two input variables jointly point to a state, their joint weight should be taken into account as well and is measured by the joint conditional probability that state θ is true given that the variables both point to it. Let $w_{\theta lm}$ denote the weight for both variables x_l and x_m that is defined as the following conditional probability that state θ is true given that state A is pointed to given e_{li} and state B is pointed to given e_{mj} with $A \cap B = \theta$:

$$w_{\theta lm} = p\left(\theta \middle| (e_{Alm}|e_{li}) \cap \left(e_{Blm}|e_{mj}\right)\right) \qquad \text{with } A \cap B = \theta \tag{50}$$

We can now define a weight index for characterising the interrelationships between the weights of two input variables.

Definition 11. (*Weight index*) The weight index of two input variables x_l and x_m for any $A, B \subseteq \Theta$ with $A \cap B = \theta$ is defined as follows:

$$\overline{w}_{Al,Bm} = p\left(\theta \middle| (e_{Alm}|e_{li}) \cap \left(e_{Blm}|e_{mj}\right)\right) \Big/ p((A|e_{Al})p(B|e_{Bm})) = \frac{w_{\theta lm}}{w_{Al}w_{Bm}} \tag{51}$$

As discussed before, p_{lm} is an evidence probability function constructed in the common *SIPS* of the data source where variables x_l and x_m take values, whilst p is a basic probability function constructed in *SOPS*. Any evidence acquired in the common *SIPS* needs to be projected to *SOPS*. This projection should follow the likelihood principle [5], so that the evidential meaning of p_{lm} is kept intact during the projection. In other words, the probability that e_{li} points to A and e_{mj} points to B measured in *SOPS*, or $p\left((e_{Alm}|e_{li}) \cap \left(e_{Blm}|e_{mj}\right)\right)$, should be proportional to the probability that state A is pointed to given $x_l = x_{li}$ and state B is pointed to given $x_m = x_{mj}$ measured in the common *SIPS* as defined in Equation (47), that is

$$p\left((e_{Alm}|e_{li}) \cap \left(e_{Blm}|e_{mj}\right)\right) = \omega_{li,mj} p_{lm}\left((e_{Alm}|x_l = x_{li}) \cap \left(e_{Blm}|x_m = x_{mj}\right)\right) \tag{52}$$

$\omega_{li,mj}$ is a non-negative constant projection rate that does not change with any $\forall \theta \subseteq \Theta$, and is uniquely determined as follows, so that Equation (17) of Definition 6 is satisfised:

$$\omega_{li,mj} = \frac{1}{\sum_{A\cap B=\theta,\ \theta\subseteq\Theta} p_{Ali,Bmj} w_{\theta lm} + 1 - r_{li,mj}} \tag{53}$$

In Equation (53), $\sum\limits_{A\cap B=\theta,\ \theta\subseteq\Theta} p_{Ali,Bmj} w_{\theta lm}$ is the expected weight of the evidence acquired in the common *SIPS* at $x_l = x_{li}$ and $x_m = x_{mj}$, and $r_{li,mj}$ the reliability of the evidence.

Definition 12. (*Projection index*) A projection index of two pieces of evidence e_{li} and e_{mj} acquired from two interrelated input variables x_l and x_m at $x_l = x_{li}$ and $x_m = x_{mj}$ is defined by

$$\overline{\omega}_{li,mj} = \frac{\omega_{li,mj}}{\omega_{li}\omega_{mj}} \tag{54}$$

In establishment of the *ER* rule, multiple pieces of evidence are assumed to be mutually independent. This is a special case in probabilistic inference. Another special case is when multiple pieces of evidence are mutually exclusive. These two special cases are defined both evidentially and cognitively as follows, where $s_{\theta lm} = \theta \cap e_{\theta lm}$ for any $\theta \subseteq \Theta$, for development of special rules and algorithms in the next section to facilitate inference in these special cases.

Definition 13. (*Evidential exclusiveness*) If it never occurs simultaneously that evidence e_{li} supports A and evidence e_{mj} supports B for any $A, B, \theta \subseteq \Theta$ with $A \cap B = \theta$, or

$$p\left((s_{Alm}|e_{li}) \cap \left(s_{Blm}|e_{mj}\right)\right) = 0,\ \forall A, B, \theta \subseteq \Theta \text{ with } A \cap B = \theta \tag{55}$$

it is then said that e_{li} and e_{mj} are evidentially exclusive of each other.

Definition 14. (*Cognitive exclusiveness*) Suppose state θ consists of multiple distinct parts. Suppose evidence e_{li} is not causally related to evidence e_{mj} and vice versa. If when e_{li} supports some parts of θ, e_{mj} always supports other parts of θ with the former never overlapping the latter, that is, $s_{Alm} \cap s_{Blm} \equiv \varnothing$ for any $A, B \subseteq \Theta$ with $A \cap B = \theta$, so that

$$p\left((s_{Alm}|e_{li}) \cap \left(s_{Blm}|e_{mj}\right)\right) = p\left(s_{Alm} \cap s_{Blm}|e_{li}, e_{mj}\right) = p\left(\varnothing|e_{li}, e_{mj}\right) = 0 \tag{56}$$

it is then said that e_{li} and e_{mj} are cognitively exclusive of each other.

Definition 13 states that two pieces of evidentially exclusive evidence do not support the same state simultaneously, or they are in complete conflict. On the other hand, Definition 14 means that they do support the same state having different constituent parts but each support different parts of the same state with no overlap; that is, they each provide their distinctive support for the same state without double counting. These definitions provide a basis to interpret interrelationships between two pieces of mutually exclusive evidence.

Definition 15. (*Evidential independence*) If the joint probability that evidence e_{li} supports state A and evidence e_{mj} supports state B is equal to the product of the probability that e_{li} supports A and the probability that e_{mj} supports B for any $A, B \subseteq \Theta$ with $A \cap B = \theta$, or

$$p\left((s_{Alm}|e_{li}) \cap \left(s_{Blm}|e_{mj}\right)\right) = p(s_{Al}|e_{li})p\left(s_{Bm}|e_{mj}\right) \tag{57}$$

it is said that e_{li} and e_{mj} are evidentially independent of each other.

Definition 16. (*Cognitive independence*) If the probability that evidence e_{li} supports state A does not depend upon whether evidence e_{mj} supports state B or not, and vice versa for any $A, B \subseteq \Theta$ with $A \cap B = \theta$, or

$$p\left(s_{Alm}|e_{li}|\left(s_{Blm}|e_{mj}\right)\right) = p(s_{Al}|e_{li}) \ \& \ p\left(s_{Blm}|e_{mj}|\left(s_{Alm}|e_{li}\right)\right) = p\left(s_{Bl}|e_{mj}\right) \tag{58}$$

it is said that e_{li} and e_{mj} are cognitively independent of each other.

Definition 15 states that two pieces of evidentially independent evidence each play their own parts in support for a state, proportional to the probability that each of them supports the state. Definition 16 means that two pieces of evidence are related to each other but the probability that one supports a state does not depend on whether the other's support is known or not.

The above-defined evidential exclusiveness (independence) is mathematically equivalent to cognitive exclusiveness (independence), but they can be interpreted differently and thus have distinctive advantages in different applications. The former may be best applied in pure data-driven inference, while the latter is better used to support knowledge-based inference.

4.2. General MAKER rule

The sort of inference investigated in this section is conjunctive in the sense that a system output is generated by combining all pieces of evidence that collectively support the states of a system space. This in essence is a process for generating the conditional probability that a state is true given that all pieces of evidence are acquired and combined. In principle, this conditional probability for any state should be proportional to the joint probability that the state is true while all pieces of evidence support it. This is one of the principles governing probabilistic inference in *ERPS*, coined as *ER* principle and explored in this paper. This principle is the *ERPS* counterpart of the widely recognised Bayesian principle governing probabilistic inference in *SOPS*. In fact, the *ER* principle in essence is the same as Bayesian principle, yet with the latter stated in *SOPS* and the former equivalently stated in *ERPS*. The *ER* principle will be strictly followed to establish probabilistic inference rules for combination of multiple pieces of evidence that are dependent on each other.

Given the above discussions, we can now establish a general conjunctive *MAKER* rule to combine two pieces of dependent evidence that jointly support a state as follows.

Theorem 1. (*MAKER rule*) The basic probability that state θ is true given both evidence e_{li} and evidence e_{mj} is the following conditional probability:

$$p\left(\theta|e_{li} \wedge e_{mj}\right) = \begin{cases} 0 & \theta = \varnothing \\ \widehat{m}_{\theta(li,mj)} \Big/ \sum_{C \subseteq \Theta} \widehat{m}_{C(li,mj)} & \theta \subseteq \Theta \end{cases} \tag{59}$$

where $\widehat{m}_{\theta(li,mj)}$ is the joint probability mass that both e_{li} and e_{mj} support θ, given by

$$\widehat{m}_{\theta(li,mj)} = \left[\left(1 - r_{mj}\right)m_{\theta li} + (1 - r_{li})m_{\theta mj}\right] + \overline{\omega}_{li,mj} \sum_{A,B \subseteq \Theta, A \cap B = \theta} \overline{w}_{Al,Bm}\overline{\alpha}_{Ali,Bmj}m_{Ali}m_{Bmj} \tag{60}$$

The trustworthiness of the combined results is given by $t_{li,mj} = \left(1 - \widetilde{m}_{2^\Theta(li,mj)}\right)$ as follows:

$$\widetilde{m}_{2^\Theta(li,mj)} = (1 - r_{li})\left(1 - r_{mj}\right) \Big/ \left(\sum_{C \subseteq \Theta} \widehat{m}_{C(li,mj)} + (1 - r_{li})\left(1 - r_{mj}\right)\right) \tag{61}$$

Proof. See Appendix A1.

The proof of Theorem 1 creates a three-step process to generate $p\left(\theta|e_{li} \wedge e_{mj}\right)$. The first step is to calculate the joint probability that both e_{li} and e_{mj} support state θ for any $\theta \subseteq \Theta$ and the joint probability of the support for the powerset, or the residual support of both e_{li}

and e_{mj} for any specific state. The second step is to calculate the joint conditional probability $P(\theta|e_{li} \wedge e_{mj})$ in *ERPS* that θ is supported for any $\theta \subseteq \Theta$ and the joint conditional probability of the support for the powerset, given both e_{li} and e_{mj}. These first two steps constitute the *ER* process and are conducted in *ERPS*. The third step is to generate $p(\theta|e_{li} \wedge e_{mj})$ in *SOPS*, which is the joint conditional probability that state θ is true given both e_{li} and e_{mj}. While $p(\theta|e_{li} \wedge e_{mj})$ is measured in *SOPS*, it should hold the same evidential meanings just as what $P(\theta|e_{li} \wedge e_{mj})$ holds in *ERPS*. In other words, there should always be $p(\theta|e_{li} \wedge e_{mj}) = k_1 P(\theta|e_{li} \wedge e_{mj})$ with k_1 being a positive constant that does not change for any $\theta \subseteq \Theta$. Theorem 1 therefore ensures that the *ER* process of conjunctively combining two pieces of evidence constitutes an augmented probabilistic inference process.

From a system point of view, if two system input variables x_l and x_m take specific values, e.g. $x_l = x_{li}$ and $x_m = x_{mj}$ with each represented by a probability distribution as defined in Equation (8), the system output generated, e.g. $y(x_l = x_{li}, x_m = x_{mj}) = y(e_{li} \wedge e_{mj})$, is also a probability distribution, as defined in Equation (2) with $p(\theta|e) = p(\theta|e_{li} \wedge e_{mj})$ that is given by Equation (59) and Equation (60). That is, Theorem 1 in essence establishes a probabilistic relationship between output y and inputs x_l and x_m. In the rest of this section, we explore the main features of Theorem 1 for probabilistic inference in several special cases.

4.3. Exclusive MAKER rule

One special case of the general *MAKER* rule is how it can be used to generate the basic probability that a state is true given two pieces of evidence that are exclusive of each other. The following corollary is the result for this special case.

Corollary 1.1. (*Exclusive MAKER rule*) If evidence e_{li} and evidence e_{mj} are exclusive of each other, Theorem 1 reduces to the following additive operation:

$$p(\theta|e_{li} \wedge e_{mj}) = \begin{cases} 0 & \theta = \emptyset \\ \dfrac{(1-r_{mj})w_{\theta l}p_{\theta li} + (1-r_{li})w_{\theta m}p_{\theta mj}}{(1-r_{mj})\sum_{C\subseteq\Theta} m_{Cli} + (1-r_{li})\sum_{D\subseteq\Theta} m_{Dmj}} & \theta \subseteq \Theta \end{cases} \tag{62}$$

Proof. See Appendix A2.

Equation (62) shows that $p(\theta|e_{li} \wedge e_{mj})$ is the mix-weighted sum of $p_{\theta li}$ and $p_{\theta mj}$ if e_{li} and e_{mj} are exclusive of each other, with the weighting factor for each piece of evidence being positively related to its own weight times the unreliability of the other evidence and normalised by a mix of the expected weight of one piece of evidence times the other's unreliability. This proves why the conjunctive combination of two pieces of exclusive evidence should also be additive. Note that a special case of Corollary 1.1 is when the weights for both pieces of evidence remain constant for any state, leading to the following conjunctive exclusive *ER* rule.

Corollary 1.2. (*Exclusive ER rule*) In Corollary 1.1, if the weights remain constant for all states, or $w_{\theta l} = w_l$ and $w_{\theta m} = w_m$ for any $\theta \subseteq \Theta$, as originally assumed in the *ER* rule, Theorem 1 reduces to the following linear operation:

$$p(\theta|e_{li} \wedge e_{mj}) = \begin{cases} 0 & \theta = \emptyset \\ \dfrac{(1-r_{mj})w_l}{(1-r_{mj})w_l + (1-r_{li})w_m}p_{\theta li} + \dfrac{(1-r_{li})w_m}{(1-r_{mj})w_l + (1-r_{li})w_m}p_{\theta mj} & \theta \subseteq \Theta \end{cases} \tag{63}$$

Proof. See Appendix A3.

Corollary 1.2 can be simplified to the following two special yet well-known operations. First, if it is further assumed that the two pieces of evidence are equally reliable, or $r_{li} = r_{mj}$, Theorem 1 reduces to the following weighted sum operation:

$$p(\theta|e_{li} \wedge e_{mj}) = \frac{w_l}{w_l + w_m}p_{\theta li} + \frac{w_m}{w_l + w_m}p_{\theta mj} \qquad \forall\theta \subseteq \Theta \tag{64}$$

If it is further assumed that the weights for both pieces of evidence are equal as well, or $w_l = w_m$, Theorem 1 reduces to the following average operation:

$$p(\theta|e_{li} \wedge e_{mj}) = \frac{1}{2}\left(p_{\theta li} + p_{\theta mj}\right) \qquad \forall\theta \subseteq \Theta \tag{65}$$

Corollary 1.1 holds for conjunctive inference when at least one piece of evidence is not fully reliable. Otherwise, both the nominator and denominator of Equation (62) will be zero, so that $p(\theta|e_{li} \wedge e_{mj})$ will be undefined. The question is how to calculate conditional probability $p(\theta|e_{li} \wedge e_{mj})$ in this special case. The following corollary provides the answer.

Corollary 1.3. (*Differential MAKER rule*) If evidence e_{li} and evidence e_{mj} are exclusive of each other and each fully reliable, or $r_{li} = r_{mj} = 1$, Theorem 1 reduces to

$$p(\theta|e_{li} \wedge e_{mj}) = \begin{cases} 0 & \theta = \emptyset \\ \dfrac{r'_{mj}w_{\theta l}p_{\theta li} + r'_{li}w_{\theta m}p_{\theta mj}}{r'_{mj}\sum_{C\subseteq\Theta} m_{Cli} + r'_{li}\sum_{D\subseteq\Theta} m_{Dmj}} & \theta \subseteq \Theta \end{cases} \tag{66}$$

r'_{li} and r'_{mj} are the derivatives of the reliabilities of e_{li} and e_{mj} at $r_{li}=1$ and $r_{mj}=1$.

Proof. See Appendix A4.

4.4. Independent MAKER rule

Another special case of the general *MAKER* rule is when it is used to generate the joint conditional probability that a state is true given two pieces of independent evidence. We have the following results to describe this special case.

Corollary 1.4. (*Independent MAKER rule*) If evidence e_{li} and evidence e_{mj} are mutually independent, Equation (60) of Theorem 1 reduces to

$$\widehat{m}_{\theta(li,mj)} = \left[(1-r_{mj})m_{\theta li} + (1-r_{li})m_{\theta mj}\right] + \sum_{A,B\subseteq\Theta,A\cap B=\theta} m_{Ali}m_{Bmj} \tag{67}$$

Proof. See Appendix A5.

In Equation (67), the first square bracket term is the sum of the probability mass that evidence e_{li} supports state θ, bounded by the unreliability of evidence e_{mj}, and the probability mass that e_{mj} supports state θ, bounded by the unreliability of e_{li}. This term is therefore referred to as the bounded sum of the probability masses of the individual support for θ from e_{li} and e_{mj}. The second term of Equation (67) is the orthogonal sum of the probability masses of the joint support for θ from e_{li} and e_{mj}. Corollary 1.4 thus asserts that the conditional probability that a state is true given two pieces of independent evidence is proportional to the addition of the bounded sum of the probability masses of their individual support and the orthogonal sum of the probability masses of their joint support for the state.

The above result is consistent with what is revealed by the *ER* rule [71]. In fact, the *ER* rule is a special case of Corollary 1.4 as follows.

Corollary 1.5. (*ER rule*) In Corollary 1.4, if it is further assumed that the weight of each input variable remains equal for all states, that is $w_{\theta l}=w_l$ and $w_{\theta m}=w_m$ for any $\theta\subseteq\Theta$, the independent *MAKER* rule given by Equation (67) reduces to the *ER* rule.

Proof. See Appendix A6.

It is in the *ER* rule that weight and reliability were originally suggested as distinctive means to model the limited importance of input variables and the inherent unreliability of evidence on the basis of rational thinking and practical experiences gained from applications in many areas for reasoning and decision making. Corollary 1.5 shows that this suggestion is deduced from the original thinking and theoretical analysis of imperfect data on the basis of both the principle of likelihood and the principle of Bayesian inference. It might seem coincidental at a first look but is in essence inevitable that these two schools of thinking give rise to the same suggestion.

Corollary 1.5 clarifies the conditions under which the *ER* rule can be applied. That is, if the support for any state from one piece of evidence is independent of the support from the other piece of evidence, then the *ER* rule can be applied to combine the two pieces of evidence, given that the weight for each piece of evidence remains equal for all states.

Dempster's rule is proven to be a special case of the *ER* rule and a reliability perturbation analysis was used to address the concerns that Dempster's rule and Bayes' rule may lead to a so-called counterintuitive problem when they are used to combine evidence in high conflict and cannot be applied to combine evidence in complete conflict [71]. The following corollary provides a theoretical and most robust answer to clarify the concerns.

Corollary 1.6. (*Augmented Dempster's rule*) In Corollary 1.5, if it is further assumed that two pieces of evidence are each fully reliable, that is $r_{li}=r_{mj}=1$, then the *ER* rule reduces to Dempster's rule, which is given as follows.

$$p(\theta|e_{li}\wedge e_{mj}) = \begin{cases} 0 & \theta=\emptyset \\ \dfrac{\sum_{A\cap B=\theta} p_{Ali}p_{Bmj}}{1-\sum_{A\cap B=\phi} p_{Ali}p_{Bmj}} & \theta\subseteq\Theta, \sigma>\tau \\ \dfrac{\overline{m}_\theta(\sigma)}{\sum_{D\subseteq\Theta}\overline{m}_D(\sigma)} & \theta\subseteq\Theta, \tau\geq\sigma>0 \\ \dfrac{\delta_{mj}w_l}{\delta_{mj}w_l+\delta_{li}w_m}p_{\theta li} + \dfrac{\delta_{li}w_m}{\delta_{mj}w_l+\delta_{li}w_m}p_{\theta mj} & \theta\subseteq\Theta, \sigma=0 \end{cases} \tag{68}$$

where $\sigma=\sum_{A\cap B=D,D\subseteq\Theta} p_{Ali}p_{Bmj}$ is the orthogonal sum of the probabilities of both e_{li} and e_{mj} supporting all states, τ is a sufficiently small probability that a state is deemed to be unlikely to occur (e.g. $\tau=0.0001$ or smaller), and $\overline{m}_\theta(\sigma)$ is the joint probability mass for θ, given by

$$\overline{m}_{\theta}(\sigma)=\left[\frac{1-r_{mj}(\sigma)}{w_m}p_{\theta li}+\frac{1-r_{li}(\sigma)}{w_l}p_{\theta mj}\right]+\sum_{A,B\subseteq\Theta,A\cap B=\theta}p_{Ali}p_{Bmj} \tag{69}$$

$r_{li}(\sigma)$ and $r_{mj}(\sigma)$ are the reliability perturbation functions for e_{li} and e_{mj}, which are valid in $\tau\geq\sigma\geq 0$. If they are required to meet the following conditions, $r_{li}(\sigma)=r_{mj}(\sigma)=1$ at σ=τ and σ=0, and $1\geq r_{li}\geq 1-\delta_{li}$ and $1\geq r_{mj}\geq 1-\delta_{mj}$ with $\delta_{li},\ \delta_{mj}\ll\tau$, then $r_{li}(\sigma)$ and $r_{mj}(\sigma)$ can in general be constructed as the following quadratic functions:

$$r_{li}(\sigma)=1-(4\delta_{li}/\tau)\sigma+\left(4\delta_{li}/\tau^2\right)\sigma^2 \text{ and } r_{mj}(\sigma)=1-\left(4\delta_{mj}/\tau\right)\sigma+\left(4\delta_{mj}/\tau^2\right)\sigma^2$$

Proof. See Appendix A7.

Note that the third and fourth rows in Equation (68) of Corollary 1.6 are presented by a general quadratic reliability perturbation function constructed for each piece of evidence. If there is no reason to believe that reliability should be perturbed differently from one piece of evidence to another, or $\delta_{li}=\delta_{mj}$, the fourth equation of Corollary 1.6 becomes a weighted sum operation. In this context, it can be asserted that Corollary 1.6 augments Dempster's rule and underpins the process of conjunctive combination of two pieces of independent and fully reliable evidence.

4.5. *MAKER algorithms*

Theorem 1 establishes the probabilistic process for conjunctive combination of two pieces of evidence. If there are more than two pieces of evidence, the process for their conjunctive combination is governed by the following Lemma 2.1 and Theorem 2.

Lemma 2.1. (*Recursive MAKER algorithm*) Suppose there are L pieces of evidence, with $e_{1\overline{\wedge}l}=(e_1\wedge\cdots\wedge e_l)$ standing for the conjunction of the first l pieces of evidence. Let $l_1=l-1$. The basic probability that evidence $e_{1\overline{\wedge}l}$ supports θ, measured in *ERPS*, is given as follows.

$$\widetilde{m}_{\theta(1\overline{\wedge}l)}=P(\theta|e_{1\overline{\wedge}l})=\begin{cases}0 & \theta=\emptyset\\ \dfrac{\widehat{m}_{\theta(1\overline{\wedge}l)}}{\sum_{C\subseteq\Theta}\widehat{m}_{C(1\overline{\wedge}l)}+\widehat{m}_{2^\Theta(1\overline{\wedge}l)}} & \theta\neq\emptyset\end{cases} \tag{70}$$

where $\widehat{m}_{\theta(1\overline{\wedge}l)}$ and $\widehat{m}_{2^\Theta(1\overline{\wedge}l)}$ are given by

$$\widehat{m}_{\theta(1\overline{\wedge}l)}=[\widetilde{m}_{\theta(1\overline{\wedge}l_1)}(1-r_l)+\widetilde{m}_{2^\Theta(1\overline{\wedge}l_1)}m_{\theta l}]+\overline{\omega}_{1\overline{\wedge}l_1,l}\sum_{A,B\subseteq\Theta,A\cap B=\theta}\overline{w}_{A(1\overline{\wedge}l_1),Bl}\overline{\alpha}_{A(1\overline{\wedge}l_1),Bl}\widetilde{m}_{A(1\overline{\wedge}l_1)}m_{Bl},\ \forall\theta\subseteq\Theta \tag{71}$$

$$\widehat{m}_{2^\Theta(1\overline{\wedge}l)}=\widetilde{m}_{2^\Theta(1\overline{\wedge}l_1)}(1-r_l) \tag{72}$$

with $\widetilde{m}_{\theta 1}=m_{\theta 1}$, $\widetilde{m}_{2^\Theta 1}=(1-r_1)$. $\overline{\alpha}_{A(1\overline{\wedge}l_1),Bl}$, $\overline{w}_{A(1\overline{\wedge}l_1),Bl}$ and $\overline{\omega}_{1\overline{\wedge}l_1,l}$ are defined in the same way as in Equations (49), (51) and (54) with e_{li} and e_{mj} replaced by $e_{1\overline{\wedge}l_1}$ and e_l, respectively.

Proof. See Appendix A8.

Lemma 2.1 describes a general recursive algorithm to combine multiple pieces of evidence in *ERPS*. In the algorithm, each piece of evidence is combined accumulatively with the previously combined evidence at each recursive step that is decomposed into two sub-steps. The first sub-step is to find the joint probability that both pieces of evidence support a state; the second sub-step is to find the conditional probability that a state is supported given both pieces of evidence. The combined evidence is then treated as a new piece of evidence that is in turn combined with another piece of evidence. This recursive process continues until all evidence is combined. All parameters for each piece of evidence are weight and reliability assigned by using domain knowledge and experiences, or trained from data, or a mix of both.

After the conditional probabilities of all states and the power set are generated by combining all pieces of evidence in *ERPS*, they are projected to *SOPS* as follows.

Theorem 2. (*Recursive MAKER rule*) The basic probability that state θ is true, given the acquisition of all L pieces of evidence $e_{1\overline{\wedge}L}=(e_1\wedge\cdots\wedge e_L)$, are generated as the following conditional probability $p(\theta|e_{1\overline{\wedge}L})$:

$$p(\theta|e_{1\overline{\wedge}L})=\begin{cases}0 & \theta=\emptyset\\ \dfrac{\widehat{m}_{\theta(1\overline{\wedge}L)}}{\sum_{C\subseteq\Theta}\widehat{m}_{C(1\overline{\wedge}L)}} & \theta\subseteq\Theta\end{cases} \tag{73}$$

where $\widehat{m}_{\theta(1\overline{\wedge}L)}$ is given by Equation (71) at l=L. The trustworthiness of the above combined results is calculated by $t_{1\overline{\wedge}L}=(1-\widetilde{m}_{2^\Theta(1\overline{\wedge}L)})$ with $\widetilde{m}_{2^\Theta(1\overline{\wedge}L)}$ given by Equation (70) for $\theta=2^\Theta$ and Equation (72) at l=L.

Proof. See Appendix A9.

Lemma 2.1 and Theorem 2 together establish the general recursive *MAKER* algorithm, to generate system output from all L inputs in a probabilistic manner in two steps. First, all pieces of evidence are represented by Definition 9 and combined recursively by applying Lemma 2.1 in *ERPS*. In the second step, the finally combined probability is projected to *SOPS* by Theorem 2, to generate system output $y(e)$ as represented by Definition 2, with $p(\theta|e)$ replaced by $p(\theta|e_{1\overline{\wedge}L})$ and the trustworthiness of the system output given by $t_{1\overline{\wedge}L} = 1 - \widetilde{m}_{2^\Theta(1\overline{\wedge}L)}$.

If multiple pieces of evidence are mutually exclusive, they can be combined by using the following analytical algorithm.

Corollary 2.1. (*Additive MAKER algorithm*) If all pieces of evidence are exclusive of each other, Lemma 2.1 and Theorem 2 reduce to the following additive algorithm.

$$p(\theta|e_{1\overline{\wedge}L}) = \begin{cases} 0 & \theta = \emptyset \\ k_a \sum_{i=1}^{L} \left(\prod_{j=1, j\neq i}^{L} (1-r_j) \right) m_{\theta i} & \theta \subseteq \Theta \end{cases} \quad k_a = \frac{1}{\sum_{C\subseteq\Theta}\sum_{i=1}^{L}\left(\prod_{j=1,j\neq i}^{L}(1-r_j)\right)m_{Ci}} \tag{74}$$

The trustworthiness of the combined results is given by $t_{1\overline{\wedge}L} = (1 - \widetilde{m}_{2^\Theta(1\overline{\wedge}L)})$ as follows:

$$\widetilde{m}_{2^\Theta(1\overline{\wedge}L)} = \frac{\prod_{j=1}^{L}(1-r_j)}{\sum_{C\in\Theta}\sum_{i=1}^{L}\left(\prod_{j=1,j\neq i}^{L}(1-r_j)\right)m_{Ci} + \prod_{j=1}^{L}(1-r_j)} \tag{75}$$

Proof. See Appendix A10.

Corollary 2.1 shows that the analytical algorithm for conjunctive combination of multiple pieces of mutually exclusive evidence is a simple additive operation. Note that $m_{\theta i} = w_{\theta i}p_{\theta i}$. The weighting factor for the i^{th} piece of evidence in the additive algorithm is thus proportional to the product of its own weight $w_{\theta i}$ and the un-reliabilities $(1 - r_j)$ of all other evidence except for its own unreliability. One special feature of this algorithm is its dictatorial behaviour in that a piece of evidence will dominate the combination if it is fully reliable, or its unreliability is zero because in this case the weighting factors for all other evidence are zero.

If multiple pieces of evidence are independent of each other, they can be combined by using the following recursive algorithm.

Corollary 2.2. (*Recursive Independent MAKER algorithm*) If all pieces of evidence are independent of each other, Equation (70) and Equation (72) in Lemma 2.1 remain unchanged but Equation (71) reduces to the following equation,

$$\widehat{m}_{\theta(1\overline{\wedge}l)} = [\widetilde{m}_{\theta(1\overline{\wedge}l_1)}(1-r_l) + \widetilde{m}_{2^\Theta(1\overline{\wedge}l_1)}m_{\theta l}] + \sum_{A\cap B=\theta, A,\ B\subseteq\Theta} \widetilde{m}_{A(1\overline{\wedge}l_1)}m_{Bl} \tag{76}$$

Proof. See Appendix A11.

In Corollary 2.2, since all pieces of evidence are mutually independent, no interdependence index needs to be calculated and only their individual weights and reliabilities need to be taken into account in inference. This makes the algorithm easy to apply in situations where subjective human judgments are needed for probabilistic inference, such as multiple criteria decision analysis under uncertainty.

The recursive *ER* rule presented by Corollary 4 in [71] is a special case of Corollary 2.2 when weight for each piece of evidence remains unchanged, or $w_{\theta l} = w_l$ for any $\theta \subseteq \Theta$ and all $l = 1, \cdots, L$. Another special case of Corollary 2.2 is that basic probabilities are assigned to singleton states and the system space only but not to any other subset of singleton states. In this case, we get the following multiplicative *MAKER* algorithm.

Corollary 2.3. (*Multiplicative MAKER algorithm*) If L pieces of evidence are independent of each other and basic probabilities are assigned to singleton states and the system space only, Lemma 2.1 and Theorem 2 reduce to the following multiplicative algorithm.

$$p(\theta|e_{1\overline{\wedge}L}) = \begin{cases} k_m\left(\prod_{i=1}^{L}(m_{\theta i} + m_{\Theta i} + (1-r_i)) - \prod_{i=1}^{L}(m_{\Theta i} + (1-r_i))\right) & \theta \in \Theta \\ k_m\left(\prod_{i=1}^{L}(m_{\Theta i} + (1-r_i)) - \prod_{i=1}^{L}(1-r_i)\right) & \theta = \Theta \\ 0 & \textit{otherwise} \end{cases} \quad k_m$$
$$= \frac{1}{\sum_{c\in\Theta}\prod_{i=1}^{L}(m_{ci} + m_{\Theta i} + (1-r_i)) - (N-1)\prod_{i=1}^{L}(m_{\Theta i} + (1-r_i)) - \prod_{i=1}^{L}(1-r_i)} \tag{77}$$

The trustworthiness of the combined results is given by $t_{1\overline{\wedge}L} = (1 - \widetilde{m}_{2^\Theta(1\overline{\wedge}L)})$ as follows:

$$\widetilde{m}_{2^\Theta(1\overline{\wedge}L)} = \frac{\prod_{i=1}^{L}(1-r_i)}{\sum_{C\in\Theta}\prod_{i=1}^{L}(m_{Ci} + m_{\Theta i} + (1-r_i)) - (N-1)\prod_{i=1}^{L}(m_{\Theta i} + (1-r_i))} \tag{78}$$

Proof. See Appendix A12.

The analytical *ER* algorithm [56] is a special case of the above multiplicative *MAKER* algorithm in that the former assumes that weight for evidence remains unchanged for all states and is equal to its reliability, or $w_{\theta l} = w_l = r_l$ for any $\theta \subseteq \Theta$ and all $l = 1, \cdots, L$. The multiplicative *MAKER* algorithm can be used to power interpretable machine learning where probabilistic inference needs to be applied.

4.6. MAKER rule and Bayesian inference

While the previous subsections focused on establishing the *MAKER* rule, its special cases and recursive and analytical algorithms, a question of particular interest is what the relationship is between the *MAKER* rule and Bayes' rule. The following corollary answers this question.

Corollary 2.4. (*Bayesian inference*) Suppose evidence e_{li} and evidence e_{mj} are the most important and fully reliable, individually or jointly, that is $w_{\theta li} = w_{\theta mj} = w_{\theta(li,mj)} = 1$ for any $\theta \subseteq \Theta$ and $r_{li} = r_{mj} = r_{li,mj} = 1$; basic probabilities for e_{li} or e_{mj} are assigned to singleton states only but not to any of their subsets, or $p_{\theta li} = p_{\theta mj} = 0$ for any $\theta \subseteq \Theta$ but $\theta \notin \Theta$. Let $c_{\theta li}$ ($c_{\theta mj}$) be the likelihood that given state θ evidence e_{li} (e_{mj}) is acquired, $c_{\theta(li,mj)}$ the joint likelihood that given state θ evidence $e_{li,mj}$ is acquired, and e_0 the prior distribution acquired independently of e_{li} and e_{mj}, which has no ambiguity either and is the most important and fully reliable as well, with $p_{\theta 0}$ being the prior probability that state θ is true. Then, the inference governed by Theorem 1 reduces to Bayesian inference as follows:

$$p\left(\theta \middle| \left(e_{li} \wedge e_{mj}\right) \wedge e_0\right) = \begin{cases} \dfrac{c_{\theta(li,mj)} p_{\theta 0}}{\sum_{A \in \Theta} c_{A(li,mj)} p_{A0}} & \theta \in \Theta \\ 0 \ \textit{otherwise} \end{cases} \tag{79}$$

If e_{li} and e_{mj} are also independent of each other, Equation (79) reduces further to

$$p\left(\theta \middle| \left(e_{li} \wedge e_{mj}\right) \wedge e_0\right) = \begin{cases} \dfrac{c_{\theta li} c_{\theta mj} p_{\theta 0}}{\sum_{A \in \Theta} c_{Ali} c_{Amj} p_{A0}} & \theta \in \Theta \\ 0 \ \textit{otherwise} \end{cases} \qquad \forall \theta \in \Theta \tag{80}$$

Proof. See Appendix A13.

Corollary 2.4 serves for three purposes. First, it shows that the *MAKER* rule as given by Theorem 1 forms a likelihood inference process. Second, it asserts that Bayesian inference is a special case of the *MAKER* rule where data is deemed perfect in the sense that (i) any variable from which evidence is acquired is the most important, (ii) any evidence acquired when any variable takes any specific value is fully reliable, and (iii) there is no ambiguity. Finally, it helps to set default initial values to estimate evidence weights and reliabilities for general inference with imperfect data by means of optimal learning, which will be explored in the next section.

5. Maximum likelihood optimal learning model

In this section, an optimisation model is constructed to learn from data the parameters that include weights for input variables and reliabilities for evidence acquired when individual or groups of input variables take specific values. It is aimed to facilitate the application of the conjunctive *MAKER* rules and algorithms in situations where available data is imperfect. It is a bi-objective nonlinear pre-emptive minimax optimisation problem, as described below.

5.1. Parameters

In Equation (60) of Theorem 1 or Equation (71) and Equation (72) of Lemma 2.1, weights $w_{\theta l}$ ($w_{\theta(1\overline{\wedge} l)}$) defined in Equation (19) for input variables and reliabilities r_{li} ($r_{(1\overline{\wedge} l)k}$) defined in Definition 8 and Equation (21) for evidence are the parameters that need to be learnt from data, or assigned using expert knowledge, or a mix of both.

The total number of weights for L input variables ($w_{\theta l}$ and $w_{\theta(1\overline{\wedge} l)}$) is denoted by N_w,

$$N_w = (|2^\Theta| - 1)L + (|2^\Theta| - 1)(L - 1) = (|2^\Theta| - 1)(2L - 1) \tag{81}$$

Let I_l be the number of values that input variable x_l takes individually and $I_{1\overline{\wedge} l} = \prod_{j=1}^{l} I_j$ the number of values that a group of input variables $\{x_1, \cdots, x_l\}$ take collectively. The total number of reliabilities to be learnt for all pieces of evidence acquired from individual and groups of input variables (r_{li} and $r_{(1\overline{\wedge} l)k}$) is denoted by N_r,

$$N_r = \sum_{l=1}^{L} I_l + \sum_{l=2}^{L} I_{1\overline{\wedge} l} = \sum_{l=1}^{L} \left(I_l + \prod_{j=1}^{l} I_j \right) - I_1 \tag{82}$$

A vector of all parameters to be learnt is denoted by λ_l and defined as follows:

$$\lambda_l = \begin{bmatrix} w_{\theta l},\ r_{li}\ \forall\theta\subseteq\Theta,\ i\in\{1,\cdots,I_l\},\ l=1,\cdots,L; \\ w_{\theta(1\overline{\wedge}l)},\ r_{(1\overline{\wedge}l)k}\ \forall\theta\subseteq\Theta, k\in\{1,\cdots,I_{1\overline{\wedge}l}\}, l=2,\cdots,L \end{bmatrix}^T \tag{83}$$

5.2. Objectives

The optimal learning model aims to estimate all the parameters, so as to make the predicted system outputs as close to the observed ones as possible as the top priority and then the value of each parameter as close to its reference value as possible for best interpretation. The learning model therefore has two objectives of different priorities. The top priority is to minimise the deviation between predicted likelihoods of system states and the observed ones for given inputs, so that the predicted likelihood of the true state is maximised. The second priority is to estimate the value of each parameter to be as close to its reference value as possible, so that the estimated parameter value can be best interpreted. The estimated parameters are used in the *ER* process to infer the maximum likelihoods of system states for given inputs. This inference process is thus referred to as maximum likelihood evidential reasoning, or *MAKER* for short.

The general formulation of the optimal learning model is constructed as the following bi-objective nonlinear pre-emptive minimax optimisation problem:

$$min\ \{f_1(\lambda), f_2(\lambda)\} \tag{84}$$

$$s.t.\ \ \lambda\in\Omega \tag{85}$$

λ is the parameter vector, including all weights and reliabilities to be learnt as defined in Equation (83), or initially $\lambda = \lambda_l$. Ω is the set of constraints that all parameters should meet. $f_1(\lambda)$ is the first objective function that counts the maximum deviation between the observed likelihood and the predicted one on any state for every observation and is of top priority. $f_1(\lambda)$ is constructed as follows.

$$f_1(\lambda) = max\{f_{11}(\lambda), f_{12}(\lambda)\} \tag{86}$$

$$f_{11}(\lambda) = \max_{l\in\{1,\cdots,L\}}\left\{\frac{S_l}{S}\max_{i\in\{1,\cdots,I_l\}}\left\{\frac{S_{li}}{S_l}\max_{A\subseteq\Theta}\left\{\frac{S_{Ali}}{S_{li}}SD_{Ali}\right\}\right\}\right\} \tag{87}$$

$$SD_{Ali} = \frac{1}{2}\sum_{\theta\subseteq\Theta}\left(\widehat{p}_{\theta(A)li} - p(\theta|e_{li})\right)^2 \tag{88}$$

$$f_{12}(\lambda) = \max_{l\in\{2,\cdots,L\}}\left\{\frac{S_{1\overline{\wedge}l}}{S}\max_{k\in\{1,\cdots,I_{1\overline{\wedge}l}\}}\left\{\frac{S_{(1\overline{\wedge}l)k}}{S_{1\overline{\wedge}l}}\max_{A\subseteq\Theta}\left\{\frac{S_{A(1\overline{\wedge}l)k}}{S_{(1\overline{\wedge}l)k}}SD_{A(1\overline{\wedge}l)k}\right\}\right\}\right\} \tag{89}$$

$$SD_{A(1\overline{\wedge}l)k} = \frac{1}{2}\sum_{\theta\subseteq\Theta}\left(\widehat{p}_{\theta(A)(1\overline{\wedge}l)k} - p(\theta|e_{(1\overline{\wedge}l)k})\right)^2 \tag{90}$$

In $f_{11}(\lambda)$, $\widehat{p}_{\theta(A)li}$ and $p(\theta|e_{li})$ are the recorded and predicted likelihoods of state θ for a data record generated by input variable l taking its i^{th} value, where state A is observed. SD_{Ali} is the sum of squared deviations between the recorded and predicted likelihoods of all states for the data record and is normalised to $0\le SD_{Ali}\le 1$. S_{li} is the number of the data records that are generated by input variable l taking its i^{th} value, and S_{Ali} the number of these data records where state A is observed. S_l is the number of all data records generated by input variable l. S is the total number of data records used to learn weights and reliabilities. $f_{11}(\lambda)$ thus measures the maximum squared deviation between recorded and predicted likelihoods for all data records generated by all single input variables individually, and is normalised to $0\le f_{11}(\lambda)\le 1$.

In $f_{12}(\lambda)$, $\widehat{p}_{\theta(A)(1\overline{\wedge}l)k}$ and $p(\theta|e_{(1\overline{\wedge}l)k})$ are the recorded and predicted likelihoods of state θ for a data record generated by input variables 1 to l taking their k^{th} combination of values, with state A observed. $SD_{A(1\overline{\wedge}l)k}$ is the sum of squared deviations between the recorded and predicted likelihoods of all states for the data record and is normalised to $0\le SD_{A(1\overline{\wedge}l)k}\le 1$. $S_{(1\overline{\wedge}l)k}$ is the number of the data records, generated by input variables 1 to l taking their k^{th} combination of values and $S_{A(1\overline{\wedge}l)k}$ the number of these data records with state A observed. $S_{1\overline{\wedge}l}$ is the number of the data records generated by input variables 1 to l. $f_{12}(\lambda)$ thus measures the maximum squared deviation between recorded and predicted likelihoods for the data records generated by all groups of input variables collectively, and is normalised to $0\le f_{12}(\lambda)\le 1$.

The first objective is therefore to minimise $f_1(\lambda)$, which measures the maximum deviation between the observed and predicted likelihoods on any state for all data records, with $0\le f_1(\lambda)\le 1$. As such, the first objective aims to minimise the maximum deviation between the recorded and predicted likelihoods of all states for each data record, in another word to maximise the predicted likelihood of the true state for each observation.

$f_2(\lambda)$ is the second objective function, which counts the maximum deviation between the reference values and the learnt values for the weights of any individual (groups of) input variables and for the reliabilities of evidence acquired from the variables. $f_2(\lambda)$ is of lower priority than the first objective and is constructed as follows.

$$f_2(\lambda) = max\{f_{21}(\lambda), f_{22}(\lambda)\} \tag{91}$$

$$f_{21}(\lambda) = \max_{l\in\{1,\cdots,L\}}\left\{\frac{S_l}{S}\max_{i\in\{1,\cdots,I_l\}}\left\{\frac{S_{li}}{S_l}max\left\{ESD_{li}, (\hat{r}_{li} - r_{li})^2\right\}\right\}\right\} \tag{92}$$

$$ESD_{li} = \sum_{\theta\subseteq\Theta}\frac{S_{\theta li}}{S_{li}}(\hat{w}_{\theta l} - w_{\theta l})^2 \tag{93}$$

$$f_{22}(\lambda) = \max_{l\in\{2,\cdots,L\}}\left\{\frac{S_{1\overline{\wedge}l}}{S}\max_{k\in\{1,\cdots,I_{1\overline{\wedge}l}\}}\left\{\frac{S_{(1\overline{\wedge}l)k}}{S_{1\overline{\wedge}l}}max\left\{ESD_{(1\overline{\wedge}l)k}, (\hat{r}_{(1\overline{\wedge}l)k} - r_{(1\overline{\wedge}l)k})^2\right\}\right\}\right\} \tag{94}$$

$$ESD_{(1\overline{\wedge}l)k} = \sum_{\theta\subseteq\Theta}\frac{S_{\theta(1\overline{\wedge}l)k}}{S_{(1\overline{\wedge}l)k}}(\hat{w}_{\theta(1\overline{\wedge}l)} - w_{\theta(1\overline{\wedge}l)})^2 \tag{95}$$

In $f_{21}(\lambda)$, $\hat{w}_{\theta l}$ and $w_{\theta l}$ are the reference and learnt weights that state θ is true given that input variable x_l points to the state, as defined in Equation (19); $\hat{r}_{li}$ and r_{li} are the reference and learnt reliabilities of evidence e_{li} acquired at $x_l = x_{li}$, as defined in Equation (21) and Definition 8. ESD_{li} is the expected squared deviation between the reference and learnt weights for evidence e_{li}, with $0 \le ESD_{li} \le 1$. $f_{21}(\lambda)$ thus measures the maximum of the expected squared deviation between the reference and learnt weights and the squared deviation between the reference and learnt reliabilities for any evidence acquired from each input variable individually. The reference weights of a variable can be assigned by examining the ability of the variable to provide correct judgments; the reference reliability of a piece of evidence can be assigned by examining the ability of the evidence to provide a correct outcome. Such assignments are problem-specific and can be data-driven or knowledge-based or a mix of both.

In $f_{22}(\lambda)$, $\hat{w}_{\theta(1\overline{\wedge}l)}$ and $w_{\theta(1\overline{\wedge}l)}$ are the reference and learnt weights that state θ is true given that a group of input variables $\{x_1, \cdots, x_l\}$ collectively point to the state. $\hat{r}_{(1\overline{\wedge}l)k}$ and $r_{(1\overline{\wedge}l)k}$ are the reference and learnt reliabilities of evidence $e_{(1\overline{\wedge}l)k}$ acquired at $x_j = x_{ji}$ for $j = 1, \cdots, l$ and $i \in \{1, \cdots, I_j\}$, with I_j being the number of values that variable j takes. $ESD_{(1\overline{\wedge}l)k}$ is the expected squared deviation between the reference and learnt weights for evidence $e_{(1\overline{\wedge}l)k}$, with $0 \le ESD_{(1\overline{\wedge}l)k} \le 1$. $f_{22}(\lambda)$ thus measures the maximum of the expected squared deviation between the reference and learnt weights and the squared deviation between the reference and learnt reliabilities for any evidence acquired from each group of input variables collectively.

Therefore, the second objective is to minimise $f_2(\lambda)$, which measures the maximum deviation between the reference and learnt values for the weight of each input variable or group of input variables and for the reliability of evidence acquired from the input variable or the group of input variables, with $0 \le f_2(\lambda) \le 1$. The learnt weight of a variable is therefore the closest to its reference weight and the learnt reliability of evidence is the closest to its reference reliability, given that the first objective is optimised and relevant constraints are met as discussed below.

5.3. Constraints

As discussed by Yang et al. [72], statistical analysis and domain knowledge can be used to set lower and upper bounds for weights and reliabilities. From the definitions of weights and reliabilities, the following general constraints should be met to construct an initial constraint set Ω_I for the optimal learning problem.

$$\Omega_I = \left\{\lambda \left| \begin{array}{l} 0 \le \underline{w}_{\theta l} \le w_{\theta l} \le \overline{w}_{\theta l} \le 1 \ \forall\theta \subseteq \Theta, \ l = 1, \cdots, L \\ 0 \le \underline{r}_{li} \le r_{li} \le \overline{r}_{li} \le 1 \ l = 1, \cdots, L, \ i = 1, \cdots, I_l \\ 0 \le \underline{w}_{\theta(1\overline{\wedge}l)} \le w_{\theta(1\overline{\wedge}l)} \le \overline{w}_{\theta(1\overline{\wedge}l)} \le 1 \ \forall\theta \subseteq \Theta, \ l = 2, \cdots, L \\ 0 \le \underline{r}_{(1\overline{\wedge}l)k} \le r_{(1\overline{\wedge}l)k} \le \overline{r}_{(1\overline{\wedge}l)k} \le 1 \ l = 2, \cdots, L, \ k = 1, \cdots, I_{1\overline{\wedge}l} \end{array}\right.\right\} \tag{96}$$

where $\overline{w}_{\theta l}$, $\overline{r}_{li}$, $\overline{w}_{\theta(1\overline{\wedge}l)}$ and $\overline{r}_{(1\overline{\wedge}l)k}$ are the upper bounds and $\underline{w}_{\theta l}$, $\underline{r}_{li}$, $\underline{w}_{\theta(1\overline{\wedge}l)}$ and $\underline{r}_{(1\overline{\wedge}l)k}$ are the lower bounds for $w_{\theta l}$, r_{li}, $w_{\theta(1\overline{\wedge}l)}$ and $r_{(1\overline{\wedge}l)k}$, respectively. If there is no prior knowledge or data to set these bounds, an upper bound is set to 1 and a lower bound to 0 by default.

Apart from the above general constraints, problem specific constraints may be added to Ω. For example, weights for some input variables may be required to be larger or smaller than others; reliabilities for some evidence may be required to be larger or smaller than others; problem specific interrelationships among weights or reliabilities may also need to be followed.

5.4. Solution methods

The optimal learning model has two objectives with different priorities. A pre-emptive solution method is adopted to solve it. First, a top priority optimisation model is constructed to minimise the first objective function $f_1(\lambda)$ as follows.

$$\min f_1(\lambda) \tag{97}$$

$$s.t. \ \ \lambda \in \Omega = \Omega_I \tag{98}$$

Suppose the optimal solution and objective value of the top priority model are given by λ^* and $f_1^* = f_1(\lambda^*)$. The second priority optimisation model is then constructed by

$$\min f_2(\lambda) \tag{99}$$

$$s.t.\ \ \lambda \in \Omega = \Omega_l;\ f_1(\lambda) \leq f_1^* + \delta \tag{100}$$

where δ is such a sufficiently small positive real number that should make the search space of the second priority model nonempty, so as to take into account any rounding-up error incurred in the nonlinear optimisation process of solving the first priority problem.

The top priority and second priority models are each a single objective optimisation model that each has a nonlinear non-smooth objective function and an initial set of constraints, mostly in the form of upper and lower bounds. They can be solved in sequence by random search algorithms, such as the Evolutionary engine of Microsoft Excel Solver.

If it is time consuming or rather difficult to find optimal solutions for problem (97)-(98) or problem (99)-(100) using random search algorithms, conventional optimisation algorithms can be used to solve them by transforming the non-smooth problems into equivalent smooth ones as follows. First, the following additional deviation variables are defined.

$$\sigma_{11li} = \max_{A\subseteq\Theta}\left\{\frac{S_{Ali}}{S_{li}}SD_{Ali}\right\} \qquad l = 1,\cdots,L;\ i = 1,\cdots,I_l \tag{101}$$

$$\sigma_{11l} = \max_{i\in\{1,\cdots,I_l\}}\left\{\frac{S_{li}}{S_l}\sigma_{11li}\right\} \qquad l = 1,\cdots,L \tag{102}$$

$$\sigma_{11} = \max_{l\in\{1,\cdots,L\}}\left\{\frac{S_l}{S}\sigma_{11l}\right\} \tag{103}$$

$$\sigma_{12(1\overline{\wedge}l)k} = \max_{A\subseteq\Theta}\left\{\frac{S_{A(1\overline{\wedge}l)k}}{S_{(1\overline{\wedge}l)k}}SD_{A(1\overline{\wedge}l)k}\right\} \qquad l = 2,\cdots,L;\ k = 1,\cdots,I_{1\overline{\wedge}l} \tag{104}$$

$$\sigma_{12(1\overline{\wedge}l)} = \max_{k\in\{1,\cdots,I_{1\overline{\wedge}l}\}}\left\{\frac{S_{(1\overline{\wedge}l)k}}{S_{1\overline{\wedge}l}}\sigma_{12(1\overline{\wedge}l)k}\right\} \qquad l = 2,\cdots,L \tag{105}$$

$$\sigma_{12} = \max_{l\in\{2,\cdots,L\}}\left\{\frac{S_{1\overline{\wedge}l}}{S}\sigma_{12(1\overline{\wedge}l)}\right\} \tag{106}$$

$$\sigma_1 = \max\{\sigma_{11},\sigma_{12}\} \tag{107}$$

$$\sigma_{21li} = \max\left\{ESD_{li},(\widehat{r}_{li} - r_{li})^2\right\} \qquad l = 1,\cdots,L;\ i = 1,\cdots,I_l \tag{108}$$

$$\sigma_{21l} = \max_{i\in\{1,\cdots,I_l\}}\left\{\frac{S_{li}}{S_l}\sigma_{21li}\right\} \qquad l = 1,\cdots,L \tag{109}$$

$$\sigma_{21} = \max_{l\in\{1,\cdots,L\}}\left\{\frac{S_l}{S}\sigma_{21l}\right\} \tag{110}$$

$$\sigma_{22(1\overline{\wedge}l)k} = \max\left\{ESD_{(1\overline{\wedge}l)k},(\widehat{r}_{(1\overline{\wedge}l)k} - r_{(1\overline{\wedge}l)k})^2\right\} \qquad l = 2,\cdots,L;\ k = 1,\cdots,I_{1\overline{\wedge}l} \tag{111}$$

$$\sigma_{22(1\overline{\wedge}l)} = \max_{k\in\{1,\cdots,I_{1\overline{\wedge}l}\}}\left\{\frac{S_{(1\overline{\wedge}l)k}}{S_{1\overline{\wedge}l}}\sigma_{22(1\overline{\wedge}l)k}\right\} \qquad l = 2,\cdots,L \tag{112}$$

$$\sigma_{22} = \max_{l\in\{2,\cdots,L\}}\left\{\frac{S_{1\overline{\wedge}l}}{S}\sigma_{22(1\overline{\wedge}l)}\right\} \tag{113}$$

$$\sigma_2 = \max\{\sigma_{21},\sigma_{22}\} \tag{114}$$

$$\lambda_d = \begin{bmatrix}\sigma_1,\sigma_{11},\sigma_{11l},\sigma_{11li},\sigma_{21},\sigma_{21l},\sigma_{21li}, l = 2,\cdots,L, i = 1,\cdots,I_l; \\ \sigma_2,\sigma_{12},\sigma_{12(1\overline{\wedge}l)},\sigma_{12(1\overline{\wedge}l)k},\sigma_{22},\sigma_{22(1\overline{\wedge}l)},\sigma_{22(1\overline{\wedge}l)k}, l = 2,\cdots,L,\ k = 1,\cdots,I_{1\overline{\wedge}l}\end{bmatrix}^T \tag{115}$$

$$\lambda = \left[\lambda_l^T,\lambda_d^T\right]^T \tag{116}$$

The top priority model (97)-(98) can then be transformed into the following equivalent smooth problem [65].

$$\min f_1(\lambda) = \sigma_1 \tag{117}$$

$$\text{s.t.} \quad \lambda \in \Omega = \Omega_l \cup \Omega_1 \tag{118}$$

$$\Omega_1 = \left\{ \lambda \left| \begin{array}{l} \sigma_{11} \le \sigma_1, \ S_l\sigma_{11l} \le S\sigma_{11}, S_{li}\sigma_{11li} \le S_l\sigma_{11l}, S_{Ali}SD_{Ali} \le S_{li}\sigma_{11li} \ \forall A \subseteq \Theta \\ \sigma_1, \sigma_{11}, \sigma_{11l}, \sigma_{11li} \ge 0 \ l = 1, \cdots, L, i = 1, \cdots, I_l; \\ \sigma_{12} \le \sigma_1, \ S_{1\overline{\wedge}l}\sigma_{12(1\overline{\wedge}l)} \le S\sigma_{12}, S_{(1\overline{\wedge}l)k}\sigma_{12(1\overline{\wedge}l)k} \le S_{1\overline{\wedge}l}\sigma_{12(1\overline{\wedge}l)}, \\ S_{A(1\overline{\wedge}l)k}SD_{A(1\overline{\wedge}l)k} \le S_{(1\overline{\wedge}l)k}\sigma_{12(1\overline{\wedge}l)k} \ \forall A \subseteq \Theta \\ \sigma_{12}, \sigma_{12(1\overline{\wedge}l)}, \sigma_{12(1\overline{\wedge}l)k} \ge 0 \ l = 2, \cdots, L; \ k = 1, \cdots, I_{1\overline{\wedge}l} \end{array} \right. \right\} \tag{119}$$

Similarly, the lower priority model (99)-(100) can also be transformed into the following equivalent smooth problem [65].

$$\min f_2(\lambda) = \sigma_2 \tag{120}$$

$$\text{s.t.} \quad \lambda \in \Omega = \Omega_l \cup \Omega_1 \cup \Omega_2 \tag{121}$$

$$\Omega_2 = \left\{ \lambda \left| \begin{array}{l} f_1(\lambda) = \sigma_1 \le f_1^* + \delta \\ \sigma_{21} \le \sigma_2, \ S_l\sigma_{21l} \le S\sigma_{21}, S_{li}\sigma_{21li} \le S_l\sigma_{21l}, \\ ESD_{li} \le \sigma_{21li}, (\hat{r}_{li} - r_{li})^2 \le \sigma_{21li}, \\ \sigma_2, \sigma_{21}, \sigma_{21l}, \sigma_{21li} \ge 0 \ l = 1, \cdots, L, i = 1, \cdots, I_l; \\ \sigma_{22} \le \sigma_2, \ S_{1\overline{\wedge}l}\sigma_{22(1\overline{\wedge}l)} \le S\sigma_{22}, S_{(1\overline{\wedge}l)k}\sigma_{22(1\overline{\wedge}l)k} \le S_{1\overline{\wedge}l}\sigma_{22(1\overline{\wedge}l)}, \\ ESD_{(1\overline{\wedge}l)k} \le \sigma_{22(1\overline{\wedge}l)k}, (\hat{r}_{(1\overline{\wedge}l)k} - r_{(1\overline{\wedge}l)k})^2 \le \sigma_{22(1\overline{\wedge}l)k}, \\ \sigma_{22}, \sigma_{22(1\overline{\wedge}l)}, \sigma_{22(1\overline{\wedge}l)k} \ge 0 \ l = 2, \cdots, L; \ k = 1, \cdots, I_{1\overline{\wedge}l} \end{array} \right. \right\} \tag{122}$$

Problems (117)-(119) and (120)-(122) can be solved using conventional optimisation software packages, such as the GRG engine of Microsoft Excel Solver.

5.5. Properties and interpretation of the model and solution

In the above subsections, the basic components of the optimal leaning model were described at the elementary level in detail. In this subsection, the principles and properties of the learning model are explicitly and formally stated to explain how the problem is formulated and why its solutions are adequate. The following proposition states the first principle that is followed with top priority for designing the learning model.

Proposition 5.1 – Maximum likelihood inference. In formulating the learning objectives, the observed and labelled data records are taken as a gold standard. The top priority of the learning model is to identify a set of weights and reliabilities that minimise the maximum squared deviation between the observed and predicted likelihoods for all training data records, that is, to maximise the predicted likelihood of the true state for each data record, thereby empowering maximum likelihood inference.

Proposition 5.1 assures that the predicted likelihoods of system output are optimal in the sense that the weights and reliabilities used for the prediction are learnt to maximise the predicted likelihood of the true state for each data record. Proposition 5.1 also implies that the learning model should be used to generate the optimal weights and reliabilities in order to legitimise *MAKER* as an optimal probabilistic inference framework.

Apart from using observed data as a gold standard, human knowledge may also be available and should be taken into account for learning weights and reliabilities, which is the second principle for designing the model as formally stated in the following proposition.

Proposition 5.2 – Referentially interpretable learning. Human knowledge is important for inference and if available should be used to set up reference weights and reliabilities. When formulating the learning objectives, the second priority is to identify a set of optimally learnt weights and reliabilities generated by satisfying the top priority objective so that the maximum deviation between the reference and learnt weights and reliabilities is minimised for all weights and reliabilities that need to be learnt.

Proposition 5.2 assures that the optimal weights and reliabilities generated by the learning model are as interpretable as their reference values in the sense that they are optimised to be as close to their reference values as possible, coined as referentially interpretable. In the current model, human knowledge is not yet treated as gold standard. Whenever they can also be regarded as a gold standard, the learning model can be easily modified by removing the priority and instead allowing trade-offs between the two objectives.

Given the decision variables discussed in Subsection 5.1 and the initial constraints discussed in Subsection 5.3, the following proposition states the features and interpretability of solutions generated by the model as well as strategies to generate the best solution among local optima.

Proposition 5.3 – Intuitively interpretable solution. The model proposed in this section empowers maximum likelihood inference and referentially interpretable learning as stated in Propositions 5.1 and 5.2 and guarantees to generate an optimal solution that could be a local optimum. Since this model is nonlinear and nonconvex and may have multiple local optima, selecting a different starting point may result in a different local optimal solution. An optimistic or a pessimistic starting point is to select the upper or lower bound, respectively, for each weight and reliability. If different starting points converge to different local optima, the one with the best overall objective value should be taken as the final solution. It is thus important to carefully set the lower and upper bounds and other constraints for training parameters, preferably based on statistical analysis or human knowledge or a mix to ensure that any local optima found within the boundaries are not unexpected and are intuitively interpretable.

While Proposition 5.3 assures the interpretability of solutions generated by the model, it also leads to questions on the model's computational complexity and feasibility. The complexity of the model depends on the amount of training data and the interrelationships or nonlinearity among multiple pieces of evidence to be combined. The more data used for training and the more complicated the interrelationships, the more complex the computation will be. If there is a huge amount of data for training, different training strategies should be considered, such as dividing the data into appropriate groups for parallel or sequential training. On the other hand, if training data is not sufficient to construct all joint likelihood functions required, joint likelihoods can also be taken as training parameters alongside weights and reliabilities. Finally, if more constraints need to be added than those initial ones shown in Equation (96), care should be taken to identify any conflicting constraints to ensure the feasibility of the model.

6. Maximum likelihood inference for disease diagnosis

In this section, the example presented in Section 2 is used to show how the conjunctive *MAKER* rule can be implemented to enable maximum likelihood inference for more trustworthy diagnosis of disease with imperfect data than the straightforward joint likelihood analysis of the data as shown in Table 10. The process of predicting the disease from saliva test and blood test is first described in the *MAKER* framework, and an optimal learning model is then constructed to estimate the model parameters to enable maximum likelihood prediction. All the calculations are in the form of Excel sheets with the Solver model provided in the Supplementary Materials.

6.1. System modelling and interdependence analysis

In the example of Section 2, if a patient has the results of both saliva test (x_1) and blood test (x_2), the question of concern is how to predict whether the patient has the disease (H_1) or not (H_2). System output (y) for this example is disease prediction and system inputs are saliva test and blood test. The system space of this example therefore has two singleton states, denoted by $\Theta = \{H_1, H_2\}$. Since the prediction can be ambiguous, the type of *SOPS* for this example is Dempster's probability space.

The discussions of Sections 3 and 4 show that the normalised likelihoods of Table 8 acquired from saliva test results and those of Table 9 from blood test results can be used for predicting the disease. For example, from a positive saliva test result, or $x_1 = x_{11}$, evidence e_{11} is acquired, as given by the second column of Table 8 as follows:

$$e_{11} = \{(e_{H_1 1},\ 0.627), (e_{H_2 1},\ 0.0535), (e_{\Theta 1},\ 0.3195)\} \tag{123}$$

as also described by Equation (10), with $e_{H_1 1} = H_{StT}$, $e_{H_2 1} = H_{StF}$ and $e_{\Theta 1} = H_{StU}$, showing the likelihoods that a positive saliva test result points to the disease being true, false or unknown, respectively, where $e_{H_1 1}$, $e_{H_2 1}$ and $e_{\Theta 1}$ are evidential elements for saliva test.

Similarly, from a positive blood test result, or $x_2 = x_{21}$, evidence e_{21} is acquired, as given by the second column of Table 9 as follows:

$$e_{21} = \{(e_{H_1 2},\ 0.6305), (e_{H_2 2},\ 0.0333), (e_{\Theta 2},\ 0.3362)\} \tag{124}$$

with $e_{H_1 2} = H_{BtT}$, $e_{H_2 2} = H_{BtF}$ and $e_{\Theta 2} = H_{BtU}$, showing the likelihoods that a positive blood test result points to the disease being true, false or unknown, respectively, where $e_{H_1 2}$, $e_{H_2 2}$ and $e_{\Theta 2}$ are evidential elements for blood test.

If a patient takes both saliva and blood tests that each turn out to be positive, evidence e_{11} of Equation (123) and evidence e_{21} of Equation (124) can be combined to infer the likelihoods that it is true, false or unknown that the patient has the disease. However, before the two test results can be combined, their interrelationships need to be analysed as they are not independent. Table 1.1 provides complete records for 1100 patients, with each having a combination of both saliva test and blood test. For each pair of tests, Table 10 shows joint likelihood distributions about how each pair of test results points to disease being true, false or unknown. This table can also be used to predict the disease. For example, from a combination of positive saliva test result and positive blood test result, or both $x_1 = x_{11}$ and $x_2 = x_{21}$, evidence $e_{11,21}$ is acquired in the second column of Table 10, which can also be denoted by $e_{(1\overline{\wedge}2)1}$ in line with the symbols used in the optimal learning model as discussed in the previous section:

$$e_{11,21} = e_{(1\overline{\wedge}2)1} = \{(e_{H_1 12},\ 0.6728), (e_{H_2 12},\ 0.0161), (e_{\Theta 12},\ 0.3111)\} \tag{125}$$

with $e_{H_1 12} = H_{SBtT}$, $e_{H_2 12} = H_{SBtF}$ and $e_{\Theta 12} = H_{SBtU}$, which depicts the likelihoods that the combination of positive saliva test result and positive blood test result points to the disease being true, false or unknown, respectively, where $e_{H_1 12}$, $e_{H_2 12}$ and $e_{\Theta 12}$ are evidential elements for both saliva test and blood test.

However, such prediction based on Table 10 alone may not be completely trusted without any concerns. For example, the 1100 patient records are imbalanced, leading to concerns on the quality of the data for inference, as analysed in Table 7, and possibly other concerns on whether it is appropriate to use the relatively small sample size of Table 1.1 to represent the whole population and that errors may be incurred during the process of generating, collecting and processing the 1100 patient records. Nevertheless, the 1100 complete patient records provide useful information on the interrelationship between the two tests, as shown in Table 10 together with Table 8 and Table 9. The information of the tables can be used to calculate interdependence indices between the results of the two tests using Equation (49) as follows:

$$\overline{\alpha}_{\theta 1i,\theta 2j} = \frac{p_{\theta 1i,\theta 2j}}{p_{\theta 1i} p_{\theta 2j}} = \frac{p_{\theta(1i,2j)}}{p_{\theta 1i} p_{\theta 2j}} \qquad \theta \subseteq \Theta,\ i,j = 1,2 \tag{126}$$

The interdependence indices are shown in Table 11. For example, the interdependence index between positive saliva test result and positive blood test result is calculated as follows. The normalised likelihood over evidential element $e_{H_1 1}$ for positive saliva test result is $p_{\theta 11} = 0.627$ in Table 8; that over evidential element $e_{H_1 2}$ for positive blood test result is $p_{\theta 21} = 0.6305$ in Table 9; that over evidential element $e_{H_1 12}$ from a combination of positive saliva result and blood test result is $p_{\theta(11,21)} = 0.6728$ in Table 10. From Equation (126), we then have $\overline{\alpha}_{\theta 11,\theta 21} = \frac{p_{\theta(11,21)}}{p_{\theta 11} p_{\theta 21}} = 1.7018$, showing that positive saliva test result and positive blood test result are more highly interrelated than being independent of each other when they jointly point to the disease being true. The other combinations of joint saliva test results and blood test results have similar features as all interdependence indices in Table 11 are larger than 1.

The likelihoods of Table 8 to Table 10 make full use of the original data given in Table 1.1 and Table 1.2 without deletion or resampling and should all be used for prediction of the disease. However, singleton saliva test, singleton blood test, or both tests each can only play a limited role in the prediction because of their limited weights and reliabilities, which need to be assigned or learnt. The weights of singleton saliva test ($w_{\theta 1}$), singleton blood test ($w_{\theta 2}$), and both tests ($w_{\theta 12}$), which can also be denoted by $w_{\theta(1\overline{\wedge}2)}$ in line with the symbols used in the optimal learning model, or $w_{\theta 12} = w_{\theta(1\overline{\wedge}2)}$ for $\theta \subseteq \Theta$, are defined by Equation (19) as the following conditional probabilities:

$$w_{\theta 1} = p(\theta|e_{\theta 1}),\ w_{\theta 2} = p(\theta|e_{\theta 2}),\ w_{\theta 12} = w_{\theta(1\overline{\wedge}2)} = p(\theta|e_{\theta 12}) \qquad \forall \theta \subseteq \Theta \tag{127}$$

The weight indices of the two tests are then calculated by Equation (51) as follows:

$$\overline{w}_{\theta 1,\theta 2} = \frac{w_{\theta 12}}{w_{\theta 1} w_{\theta 2}}\ \forall \theta \subseteq \Theta \tag{128}$$

Similarly, there is a need to assign or learn the reliabilities of evidence given in Table 8 and illustrated in Equation (123) for saliva test results (r_{1i}), in Table 9 and Equation (124) for blood test results (r_{2j}), and in Table 10 and Equation (125) for both test results ($r_{1i,2j}$), which can also be denoted by $r_{(1\overline{\wedge}2)k}$ in line with the symbols used in the optimal learning model, or $r_{1i,2j} = r_{(1\overline{\wedge}2)k}$ with $k = j + 2(i-1)$ and $i,j = 1,2$, so $k = 1,\cdots,4$. The projection rate is defined in Equation (21) and Equation (53), i.e. ω_{1i} for saliva test results, ω_{2j} for blood test results, and $\omega_{1i,2j}$ for both test results. The projection indices of the tests are then calculated as follows:

$$\overline{\omega}_{1i,2j} = \omega_{1i,2j} / (\omega_{1i}\omega_{2j}) \quad \text{for} \quad i,j = 1,2 \tag{129}$$

6.2. Construction of optimal learning model

As discussed in the previous subsection, we need to estimate the following parameters: weights $w_{\theta 1}$, $w_{\theta 2}$, and $w_{\theta(1\overline{\wedge}2)}$ for all $\theta \subseteq \Theta$; reliabilities r_{1i}, r_{2j}, and $r_{(1\overline{\wedge}2)k}$ for $i,j = 1,2$ and $k = 1,\cdots,4$. In addition, since all the data in Table 1.1 and Table 1.2 is used to estimate these parameters, the prior of all the data need to be combined to reduce the effect of data imbalance on parameter estimation. The prior distribution of all the data is calculated as follows:

In Table 12, $e_{\theta 0}$ for $\theta \subseteq \Theta$ is the evidential element that the prior points to the disease being true ($\theta = H_1$), false ($\theta = H_2$) and unknown ($\theta = \Theta$), respectively. The weights and reliability of the prior are denoted by $w_{\theta 0}$ for $\theta \subseteq \Theta$ and r_0.

A vector of all 21 parameters to be learnt in this example is represented by λ_l as follows:

$$\lambda_l = [w_{\theta l}, w_{\theta(1\overline{\wedge}2)}, \forall \theta \subseteq \Theta, l = 0,\ \cdots, 2;\ r_0, r_{li}, r_{(1\overline{\wedge}2)k}, l, i = 1,2, k = 1,\cdots,4]^T \tag{130}$$

In the first objective function $f_1(\lambda)$ of the optimal learning model, $p(\theta|e_{li})$ in Equation (88) is the conditional probability of state θ that is generated by applying Theorem 1 to combine evidence e_{li} acquired from the i^{th} result of the l^{th} test with prior e_0 as a piece of independent evidence; $\widehat{p}_{\theta(A)li}$ in Equation (88) is the probability of state θ given in a patient record where state A is observed when the l^{th} test takes its i^{th} result for $l,i = 1,2$. From the data of Table 1.2, there are $\widehat{p}_{\theta(A)li} = 1$ at $\theta = A$ and $\widehat{p}_{\theta(A)li} = 0$ at $\theta \neq A$ for any $\theta, A \subseteq \Theta$ and $l,i = 1,2$; the frequency data defined in sub-objective functions $f_{11}(\lambda)$ are given in Table 13 and Table 14, with $S_1 = S_{11} + S_{12} = 2100$ and $S_2 = S_{21} + S_{22} = 3100$.

$p(\theta|e_{(1\overline{\wedge}2)k})$ in Equation (90) is the probability generated by applying Theorem 1 to combine evidence e_{1i} acquired from the i^{th} saliva

Table 11
Interdependence Index between Saliva and Blood Tests.

Interdependence index	Positive Saliva		Negative Saliva	
	Positive Blood	Negative Blood	Positive Blood	Negative Blood
H_1	1.7018	9.1909	15.3551	3.7689
H_2	9.0204	15.6675	11.8828	1.6542
Θ	2.8966	2.2425	2.7895	3.1465

Table 12
Prior frequency and distribution.

	$e_{H_1 0}$	$e_{H_2 0}$	$e_{\Theta 0}$	Total
Frequency	565	5435	300	6300
e_0	0.0897	0.8627	0.0476	

test result with evidence e_{2j} acquired from the j^{th} blood test result with their interdependency taken into account, which is in turn combined with prior e_0 as a piece of independent evidence for all $i, j = 1,\ 2$ and $k = 1, \cdots, 4$. Accordingly, $\widehat{p}_{\theta(A)(1\overline{\wedge}2)k}$ in Equation (90) is the probability of state θ given in a patient record generated for the k^{th} combination of the i^{th} saliva test result and the j^{th} blood test result for $k = 1, \cdots, 4$, where state A is observed. From the data of Table 1.1, there are $\widehat{p}_{\theta(A)(1\overline{\wedge}2)k} = 1$ at $\theta = A$ and $\widehat{p}_{\theta(A)(1\overline{\wedge}2)k} = 0$ at $\theta \neq A$ for any $\theta, A \subseteq \Theta$ and $k = 1, \cdots, 4$; the frequency data defined in sub-objective function $f_{12}(\lambda)$ are given in Table 15, with $S_{1\overline{\wedge}2} = S_{(1\overline{\wedge}2)1} + \cdots + S_{(1\overline{\wedge}2)4} = 1100$ and $S = S_1 + S_2 + S_{1\overline{\wedge}2} = 6300$.

In the second objective function $f_2(\lambda)$, $\widehat{w}_{\theta l}$ in Equation (93) is the reference weight of the l^{th} test for state θ, and is assumed to be the upper bound of $w_{\theta l}$ in this example, or $\widehat{w}_{\theta l} = \overline{w}_{\theta l}$ for any $\theta \subseteq \Theta$ and $l = 1, 2$, under the notion that saliva or blood test should play its most important role in inference. $\widehat{w}_{\theta(1\overline{\wedge}2)}$ in Equation (95) is the reference weight of both saliva and blood tests for state θ and is also assumed to be the upper bound, or $\widehat{w}_{\theta(1\overline{\wedge}2)} = \overline{w}_{\theta(1\overline{\wedge}2)}$ for any $\theta \subseteq \Theta$. $\widehat{r}_{li}$ in Equation (92) is the reference reliability of evidence e_{li}, and is assumed to be the upper bound of r_{li}, or $\widehat{r}_{li} = \overline{r}_{li}$ for any $l, i = 1, 2$ under the notion that any evidence should be as reliable as possible in inference. $\widehat{r}_{(1\overline{\wedge}2)k}$ in Equation (94) is the reference reliability of evidence $e_{(1\overline{\wedge}2)k}$ and is assumed to be the upper bound of $r_{(1\overline{\wedge}2)k}$, or $\widehat{r}_{(1\overline{\wedge}2)k} = \overline{r}_{(1\overline{\wedge}2)k}$ for $k = 1, \cdots, 4$.

In the optimal learning model, constraints include only upper and lower bounds for the 21 parameters to be learnt. In this study, the data used to acquire evidence is given in Table 4 for $e_{(1\overline{\wedge}2)k}$ ($k = 1, \cdots, 4$), in Table 5 for e_{1i} ($i = 1, 2$), and in Table 6 for e_{2j} ($j = 1, 2$). These three tables all originate from the same data source of Table 1.1 and Table 1.2. The quality of the data is assumed to be at least as good as completely random, so the lower bounds of all weights and reliabilities are assumed to be 0.5, that is, $\underline{w}_{\theta l} = 0.5$, $\underline{r}_{li} = 0.5$, $\underline{w}_{\theta(1\overline{\wedge}2)} = 0.5$ and $\underline{r}_{(1\overline{\wedge}2)k} = 0.5$ for $l, i = 1, 2$, $k = 1, \cdots, 4$ and all $\theta \subseteq \Theta$. Since the original data of Table 1.1 and Table 1.2 is imperfect, no variable should be assumed to be as important as 100% and any evidence acquired from the variables should not be assumed to be fully reliable either. Instead, the weights of each variable and the reliability of any acquired evidence should have upper bounds less than 1 in general. For illustration purpose, to reflect the difference of sample sizes in these tables, the upper bounds $\overline{w}_{\theta 2}$ and $\overline{r}_{2j}$ are assumed to be 0.95 as Table 6 has the largest sample size of 4200 patient records, followed by $\overline{w}_{\theta 1}$ and $\overline{r}_{1i}$ to 0.9, and $\overline{w}_{\theta(1\overline{\wedge}2)}$ and $\overline{r}_{(1\overline{\wedge}2)k}$ to 0.85 for i, $j = 1, 2$, $k = 1, \cdots, 4$ and all $\theta \subseteq \Theta$.

The optimal learning model is therefore formulated as a bi-objective nonlinear non-smooth mathematical programming problem, having 2 prioritised objectives and 21 variables with their lower and upper bounds as constraints. In this study, the Evolutionary engine of Excel Solver is selected to solve the problem.

6.3. Results and analysis

A pre-emptive solution method is used to solve the above problem. The first objective $f_1(\lambda)$ is of the top priority and is minimised first by setting the starting values of all the parameters to the upper bounds. The optimal solution is given by $f_1^* = f_1(\lambda^*) = 0.0075$. The second objective $f_2(\lambda)$ is then minimised by adding the priority constraint: $f_1(\lambda) \leq f_1^* + \delta$ with $\delta = 0.00001$. The optimal solution for the second objective is given by $f_2^* = 0.0247$. The detailed calculations can be found in the provided Supplementary Materials in the form of Excel sheets with the Solver model attached. The optimal results are analysed as follows.

The optimal values of the weights and reliabilities for saliva test, blood test and the prior are given in Table 16, where $w_{\theta l}^*$ is the optimal weight of saliva test (l=1), blood test (l=2) or the prior (l=0) for state θ. r_{li}^* is the optimal reliability of evidence e_{li} acquired from the positive (i=1) or negative result (i=2) of saliva test (l=1) or blood test (l=2). r_0^* is the optimal reliability of evidence e_0.

The optimal values of the weights and reliabilities for both saliva test and blood test are given in Table 17, where $w_{\theta(1\overline{\wedge}2)^*}$ is the optimal weight of both saliva test and blood test for state θ. $r_{(1\overline{\wedge}2)k^*}$ is the optimal reliability of evidence $e_{(1\overline{\wedge}2)k}$ acquired from the k^{th} combination of saliva test result and blood test result for $k = 1, \cdots, 4$.

One observation from the results of Table 16 and Table 17 is that the reliability of any evidence acquired from a saliva result, a blood test result or from the prior is higher than that from any combination of both saliva test result and blood test result. This is likely due to the fact that there are only 1100 patient records that have both saliva test results and blood test results, whilst 3200, 4200 and

Table 13
Frequency for Saliva Test.

Frequency	Positive	Negative
$S_{H_1 1i}$	180	15
$S_{H_2 1i}$	155	1650
$S_{\Theta 1i}$	50	50
S_{1i}	385	1715

Table 14
Frequency for Blood Test.

Frequency	Positive	Negative
$S_{H_1 2j}$	250	20
$S_{H_2 2j}$	140	2590
$S_{\Theta 2j}$	54	46
S_{2j}	444	2656

Table 15
Frequency for Both Saliva and Blood Tests.

Frequency	Saliva Positive		Saliva Negative	
	Blood Positive	Blood Negative	Blood Positive	Blood Negative
$S_{H_1(1\overline{\wedge}2)k}$	93	2	4	1
$S_{H_2(1\overline{\wedge}2)k}$	20	40	20	820
$S_{\Theta(1\overline{\wedge}2)k}$	43	2	3	52
$S_{(1\overline{\wedge}2)k}$	156	44	27	873

6300 patient records are used to acquire evidence from saliva test, blood test and the prior, respectively.

The maximum likelihoods $p(\theta|e_{1i} \wedge e_{2j})$ for $i, j = 1, 2$ and all $\theta \subseteq \Theta$ are shown in Table 18, which are generated by using the optimal values of weights and reliabilities to combine every pair of saliva and blood test results without the prior taken into account.

Comparing the maximum likelihoods of Table 18 with the ordinary likelihoods of Table 10 for all $k = 1, \cdots, 4$ and $\theta \subseteq \Theta$, one can tell that they are quite similar but also different from each other. They are similar as the former are generated from all the available 6300 patient records of both Table 1.1 and Table 1.2 and the latter from a small yet still respectable size of 1100 complete patient records of Table 1.1. They are different because of their different sample sizes and also because the data is imperfect, e.g. ambiguous, incomplete and imbalanced.

For example, if the saliva test result of a patient is positive but his blood test result is negative, Table 10 shows that it is more likely that the patient has no disease with $p_{H_{BStF},(1\overline{\wedge}2)2} = 0.5263$, but in Table 18 the corresponding maximum likelihood is $p(H_2|e_{11} \wedge e_{22}) = 0.4909$, showing a more uncertain diagnosis for the disease. A more detailed analysis of the original data reveals why such difference occurs. In Table 8, for a positive saliva test result we have $p_{H_{StT},11} = 0.6270$, mostly pointing to state H_1, but in Table 9 for a negative blood test result $p_{H_{BtF},22} = 0.6283$, mostly pointing to state H_2, so the two results are in conflict for predicting the disease. While $p_{H_{StT},11}$ is almost the same as $p_{H_{BtF},22}$, the trustworthiness of the former is 0.8861 but that of the latter is a bit lower at 0.8602. This explains why we get $p(H_2|e_{11} \wedge e_{22}) = 0.4909$ in Table 18, the trustworthiness of which is 0.9834, much higher than that for $e_{(1\overline{\wedge}2)2}$ that is 0.7223 in Table 17. The above analysis reveals the main difference between the results of Table 18 and the results of Table 10 generated from the straightforward likelihood analysis of the original data of Table 1.1. That is, the former can be more highly trusted than the latter. In fact, the prediction of Table 18 has much higher trustworthiness than that of Table 10 for every combination of saliva test results and blood test results.

Similar observations can be made between the results of Table 18 and those of Table 8 (Table 9) generated from the straightforward likelihood analysis of the original data of Table 5 (Table 6). In fact, comparing the trustworthiness of Table 16 with that of Table 18, we can see that the prediction of Table 18 also has higher trustworthiness than that of Table 8 (Table 9) whether saliva (blood) test result is positive or negative. For example, the trustworthiness of Table 18 for positive saliva test result is 0.9882 (0.9834) if blood test result is positive (negative), each higher than that of Table 8 for positive saliva test result (evidence e_{11}), which is 0.8861 as shown in the t^*_{li} row and the e_{11} column of Table 16.

The above observations are made not by chance but reveal one of the prominent features of *MAKER* in that in general the combination of multiple pieces of evidence, which are each trusted to some degrees, leads to better trusted conclusions than any individual evidence.

Table 16
Parameter Values for Saliva, Blood and the Prior.

Optimal parameters	Saliva test		Blood test		Prior
	e_{11}	e_{12}	e_{21}	e_{22}	e_0
r^*_{li}	0.8948	0.8419	0.9167	0.8728	0.8817
$w^*_{H_1 l}$	0.8503		0.9437		0.9952
$w^*_{H_2 l}$	0.7479		0.9196		0.9162
$w^*_{\Theta l}$	0.7693		0.5027		0.5000
t^*_{li}	0.8861	0.8278	0.9052	0.8602	0.8842

Table 17
Parameter Values for Both Saliva and Blood.

Optimal parameters	Saliva			
	Positive		Negative	
	Blood		Blood	
	Positive $e_{(1\overline{\wedge}2)1}$	Negative $e_{(1\overline{\wedge}2)2}$	Positive $e_{(1\overline{\wedge}2)3}$	Negative $e_{(1\overline{\wedge}2)4}$
$r_{(1\overline{\wedge}2)k^*}$	0.7297	0.7253	0.7001	0.7638
$w_{H_1(1\overline{\wedge}2)^*}$	0.6867			
$w_{H_2(1\overline{\wedge}2)^*}$	0.7237			
$w_{\Theta(1\overline{\wedge}2)^*}$	0.7217			
$t_{(1\overline{\wedge}2)k^*}$	0.7209	0.7223	0.7022	0.7537

Table 18
Maximum Likelihood Generated by Combining Saliva and Blood Tests.

Maximum Likelihood	Saliva Positive		Saliva Negative	
	Blood Positive $p(\theta\|e_{11}\wedge e_{21})$	Blood Negative $p(\theta\|e_{11}\wedge e_{22})$	Blood Positive $p(\theta\|e_{12}\wedge e_{21})$	Blood Negative $p(\theta\|e_{12}\wedge e_{22})$
H_1	*0.6706*	*0.2647*	*0.4447*	*0.0178*
H_2	*0.0221*	*0.4909*	*0.2415*	*0.6445*
Θ	*0.3072*	*0.2444*	*0.3138*	*0.3377*
$t_{1i,2j}$	0.9882	0.9834	0.9829	0.9769

7. Discussions, conclusions and directions for future research

7.1. Discussions

The *MAKER* framework established in this paper takes a system view and is underpinned by the widely recognised scientific principles governing probabilistic inference, that is, the likelihood principle and Bayesian principle as stated in *SOPS* or the *ER* principle as equivalently stated in *ERPS*. It provides a unified framework to model randomness, ambiguity and untrustworthiness for probabilistic inference and evidence-based decision making in a principled, thus intrinsically transparent and interpretable manner.

In the *MAKER* framework, randomness of a singleton state in a system space associated with a piece of evidence is described by assigning a basic probability to the singleton state, which is entirely consistent with how it is modelled in Bayesian inference. Ambiguity among a subset of singleton states associated with the evidence is depicted by assigning a basic probability to the subset as a whole, which represents the degree of unknown or ignorance about these singleton states and cannot be split into pieces to be assigned to any singleton state of the subset. This is completely in line with Dempster's principle, which ought to be followed to handle ambiguity for probabilistic inference. Untrustworthiness of the evidence is explicitly measured by retaining it uniquely to the powerset of the system space as a whole. It is originally calculated from the reliability and expected weight of the evidence, which is acquired from a data source when a specific value is taken by an input variable for which the weights are assigned.

The weight of an input variable provides an appropriate measure of its perceived ability to provide correct values or judgments. It is unambiguously defined as the conditional probability that a state is true given that the variable points to the state. The trustworthiness of evidence is defined to be positively related to the expected weight of the evidence. That is, more highly weighted evidence is more highly trusted, and vice versa. The reliability of evidence provides a clear measure of its ability to provide correct outcome in terms of probability distribution on all states. It is so defined that the unreliability of the evidence is positively related to the probability that the evidence is untrusted to support any state, which is thus referred to as residual support, kept intact and retained to the power set of the system space, ready for combination with other evidence. The untrustworthiness of evidence is positively related to the unreliability of the evidence. That is, more unreliable evidence is more untrusted, and vice versa.

The general *MAKER* rule and algorithm are established by strictly following the widely recognised scientific principles and provide a general means to enable probabilistic inference with multiple pieces of evidence that (i) can be acquired from imperfect data or human knowledge, (ii) are each associated with weight and reliability and (iii) are dependent on each other in general. In the *MAKER* framework, interdependency among multiple pieces of evidence is accurately depicted and explicitly taken into account, so that complex system behaviours can be precisely captured in inference. The special rules and algorithms deduced from the general *MAKER* rule and algorithm can be used to facilitate probabilistic inference in situations where evidence is exclusive or independent of each other. The identified and proven conditions where the *MAKER* rule reduces to the *ER* rule, Dempster's rule and Bayes' rule reinforce the robustness and flexibility of the *MAKER* rule and algorithm for probabilistic reasoning and decision making.

The bi-objective nonlinear minimax optimisation model provides a general and robust means to learn variable weights and evidence reliabilities from observation data, where imperfection in data and human knowledge as well as human preferences can be taken into account as model constraints or through adjustment of the objectives according to different scenarios where the model is applied.

The model is constructed in both non-smooth format and equivalent smooth format, so that it can be implemented in different optimisation platforms. The experiences of building and solving the model for the two illustrative examples and the case study show that it is easy and fast to implement the model in Microsoft Excel Solver. It should also be easy to implement the model in other similar or more powerful platforms.

The examination of the two examples and the case study given in Appendix B provides further insights of the *MAKER* framework and how it could be applied in many other fields than what are implied by these cases, e.g. the first example for healthcare, the second example for legal services and the case study for human well-being. Although the data for the first example is simulated from the authors' experiences gained from conducting many research projects in healthcare, it shows how typical features common to imperfect data can be modelled and handled within the *MAKER* framework without making unnecessary and unjustifiable assumptions in the modelling and inference process. The examination of the second example provides insights into how and why the *MAKER* framework can be used to help deal with complex yet serious issues in decision making in legal services and other professional services in general. The case study shows that even though the *BHPS* data is carefully collected and of high quality, it is still imperfect due to missing and imbalanced data, whilst the *MAKER* framework is equipped with the ability to handle imperfect data as it is, whether it is of high quality or not.

7.2. Conclusions

This paper presented the findings generated from the dedicated research on generalising the *ER* rule. Its main contributions to knowledge are summarised as follows.

- The most significant contribution of the paper is the establishment of the unique *MAKER* framework, which consists of three constituent models: *SSM, EAM* and *ERM* innovatively constructed in three interconnected probability spaces: *SOPS, SIPS* and *ERPS*, respectively, to facilitate probabilistic inference under three typical types of uncertainty: randomness, ambiguity and untrustworthiness. While *SSM* was introduced to implement Dempster's principle for probabilistic inference with ambiguity, *EAM* and *ERM* stem from the authors' original thinking of creating *SIPS* and *ERPS*, leading to a paradigm shift in research on probabilistic inference with the three types of uncertainty under the same umbrella.
- The establishment of *EAM* in *SIPS* and *ERM* in *ERPS* in turn leads to the paper's second most significant contribution of identifying the precise definitions of variable weight and evidence reliability, vital to enable the informed assignment of weight and reliability using human knowledge and by interpretable and optimal learning from imperfect data. The explicit definition of trustworthiness in terms of the reliability and expected weight of individual pieces of evidence and the calculation of trustworthiness for the combined evidence provide transparent and robust means to measure the degree to which an evidence-based decision or conclusion can be trusted.
- The derivation and proofs of the *MAKER* rules and algorithms are the unique contributions of the paper that not only lay the solid theoretical foundation but also provide powerful technical means for developing novel models, tools and platforms for probabilistic inference by combining multiple pieces of evidence that can be dependent in general. In particular, the additive or multiplicative *MAKER* algorithm provides a simple and powerful formula for probabilistic inference by combining multiple pieces of evidence that are exclusive or independent of each other, respectively.
- The differential *MAKER* rule lays the theoretical foundation for how Bayes' and Dempster's rules should be applied when used to combine multiple pieces of fully reliable evidence that are completely or highly conflicting. This leads to the paper's contributions of creating the theoretical and most robust resolution to the so-called counterintuitive problem associated with Dempster's rule and Bayes' rule, a long-term hassle to the community of probabilistic inference, augmenting and completing Dempster's rule, and identifying and proving the conditions where the *MAKER* rule reduces to the *ER* rule, Dempster's rule and Bayes' rule.
- The constructed bi-objective optimal learning model is another contribution that provides a transparent framework and operational means to learn optimal weights and reliabilities as well as joint probabilities, which are essential to power maximum likelihood inference when data is imperfect or insufficient to construct valid likelihood functions.
- The examined two numerical examples and the case study contribute to helping understand how the *MAKER* framework can be applied in different domains, in particular healthcare, professional services and human wellbeing analysis. It is shown in the paper how real or hypothetical data can be used to construct the bi-objective optimal learning model for generating optimal weights and reliability, how to use the learnt weights and reliabilities to assign trustworthiness for individual pieces of evidence and how to generate overall trustworthiness for a conclusion or decision as a result of combining multiple pieces of evidence each with its individual trustworthiness.

While the above main contributions are characteristic of the paper and seem natural and straightforward to summarise, the potential contributions of the paper are plenty and could be even more significant. For instance, the *MAKER* framework and constituent rules and algorithms could underpin a paradigm shift in future research on *AI*-powered decision theories and systems; they together with the optimal learning model would also help boost a step change in advancing interpretable machine learning and reinforcement learning with human interaction such as in the domains of interactive multi-objective optimisation and multi-criteria decision making.

While the *MAKER* framework provides a new perspective for modelling any linear or nonlinear relationships between system outputs and inputs under different types of uncertainty characterised by imperfect data, there are challenges that can be encountered when applying the framework. The first challenge is to identify whether a system analysis problem is a prediction or decision problem or a hybrid one. A different problem can imply different relationships between system outputs and inputs, which can be causal, correlational or mixed.

For instance, the first numerical example of disease diagnosis given medical tests is a prediction problem where the relationship between the disease and the tests is correlational rather than causal. The second numerical example of jury conviction is a decision problem where a juror's conclusion contributes to or causes the decision of the jury. The case study of human wellbeing analysis is also a decision problem where a person's health and income influence or cause her life satisfaction. Many real-world system analysis problems can be more complicated and require the construction of multiple interconnected models where the outputs of some models can be the inputs of others and their relationships can be hybrid. While *MAKER* provides a general modelling framework, domain specific knowledge and system analysis skills are needed to develop appropriate models for complex system analysis, modelling, prediction and decision making.

The second challenge is how to construct likelihood functions from imperfect data. In the first numerical example, the imperfection of data was explained and the process of applying the likelihood analysis method to construct likelihood functions was demonstrated, However, that is a rather simple example with only two input variables and each having only two categorical values. With the increase of variables and categorical or referential values, it will become more challenging to generate meaningful likelihood functions using the likelihood analysis method.

For high-dimensional problems with many input variables and limited data available, there will be a need to learn likelihood functions from limited imperfect data. The *MAKER* framework provides a transparent structure to enable interpretable machine learning of variable weights and evidence reliabilities. However, it will be another challenge to develop interconnected models of hierarchical or network structures for dealing with high-dimensional system modelling and analysis. While domain specific knowledge is useful for developing model hierarchies or networks, it poses a significant challenge to develop adaptive machine learning models that can learn optimal weights, reliabilities and probably joint likelihoods from imperfect data.

When system analysis needs to be based on human judgments, such as in the second example of jury conviction, educated assumptions need to be made in order to apply the *MAKER* framework. For instance, to apply the *ER* rule to combine multiple jurors' conclusions in the second example, every effort needs to be made to keep a juror's judgment independent of others; no juror's individual judgment should be assumed to be fully reliable; no juror should be allowed to dominate the jury's decision-making process.

It is also worth noting that the bi-objective optimisation model devised in Section 5 to learn variable weights and evidence reliabilities relies on not only observation data between inputs and outputs but also how the data is generated, collected and processed in specific problem domains such as fault diagnosis [72]. It is therefore important to use domain knowledge for estimating the upper and lower bounds of weights and reliabilities and add them as constraints to the optimal learning model so that the optimally learnt weights and reliabilities can be adequately interpreted.

7.3. Possible directions for future research

This paper aimed to report the theoretical development and advancement of the *ER* rule, More and wider theoretical and applied research is undoubtedly needed to make *AI* theories, methodologies, tools and systems as transparent, interpretable, trustworthy and thereby acceptable as possible. Possible directions for future research in probabilistic inference are briefly discussed as follows.

The two numerical examples and the case study reported in this paper were intended for illustration purpose by using real and hypothetical data and the Excel Solver platform to demonstrate how the new *MAKER* framework, rules, algorithms and optimal learning model are operated step by step. The authors have led or been involved in several research and development projects, such as those mentioned in the Acknowledgement section, where some of the *MAKER* framework and algorithms have been applied and the specially designed software platforms have been developed in the domain of professional services, in particular insurance and legal services [15]. These platforms have been developed for commercial purposes but not for supporting academic research or general applications of the *MAKER* framework.

Therefore, the immediate future research is to develop new open-source software platforms for automatically conducting the likelihood analysis of imperfect data and implementing the multiplicative and additive *MAKER* algorithms and the optimal learning model. Without such open-source platforms, it is difficult to implement the *MAKER* framework for supporting further research and wider applications in different domains, such as healthcare, professional services, sustainability, transportation systems, manufacturing and supply chain systems.

Further theoretical and methodological research is needed on how to learn variable weights and evidence reliabilities based on both the optimal learning model and domain specific knowledge and how to apply the *MAKER* framework to develop novel methods and systems for multiple objective optimisation (*MOO*) and multiple criteria decision making (*MCDM*) under uncertainty, which are among the main motivations for establishing the *ER* rule. While the *ER* rule is restricted to probabilistic inference with independent evidence, the *MAKER* framework opens plenty of new opportunities to develop new *MOO* and *MCDM* methods and systems to support evidence-based decision making in much wider ranges of domains where evidence can be either independent or exclusive or dependent in general.

The new bi-objective optimal learning model and the likelihood analysis method, together with the belief rule base methodology [68], opens a new avenue for advancing research in interpretable machine learning with both imperfect data and human knowledge. Also, future research needs to be conducted on advancing interpretable reinforcement learning with human interaction powered by the *MAKER* framework and algorithms in combination with new *MOO* and *MCDM* methods, where decision makers preferences can be learnt in an interactive fashion and used to construct utility functions to guide the search for most preferred solutions.

Last but not least, dedicated research is needed to enable disjunctive probabilistic inference with imperfect data under different types of uncertainty, in particular randomness, ambiguity and untrustworthiness. While the research may be conducted within the

MAKER framework, the derivation and proofs of disjunctive rules and algorithms will pose significant challenges given that disjunctive probabilistic inference needs to be established on top of conjunctive probabilistic inference that is the focus of this paper. Nevertheless, disjunctive probabilistic inference is widespread and essential for developing *AI* systems in many fields, such as decision making under risk, system safety analysis, fault diagnosis and medical prognosis.

CRediT authorship contribution statement

Jian-Bo Yang: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. **Dong-Ling Xu:** Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Validation, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

It would not be possible to complete this paper without the persistent efforts of so many researchers working so effectively to improve and apply the *ER* rule, including those whose papers are cited in this paper. The authors are grateful to all these researchers for their great efforts and contributions that have motivated the research from different perspectives. Special thanks to Professor Arthur Dempster, who provided invaluable advice to the authors both in person and via email. It is his pioneer work and system view that set a most important foundation for the research. It is due to his advice that a principled approach is explored to establish the *MAKER* framework. Special thanks also to Dr Lin Yang, who not only provided unique encouragement for the research and the *BHPS* data but also offered invaluable expert advice for the case study and kindly proofread the manuscript as well. The authors are grateful to the editors of the *AI* journal and the anonymous reviewers for their constructive comments and invaluable encouragement that have helped strengthen the paper to meet the high standard of the journal. In the long period of conducting the research, the authors have benefited from and been partially supported by many research projects, including the current Innovate UK project under grant No.: 10021202 (Reputation Advisor), European Union's Horizon 2020 Research and Innovation Programme RISE under grant No.: 823759 (ReMESH), the Natural Science Foundation of China under Grant No.:72010107002, and InnoHK initiative, the Government of the HKSAR, and Hong Kong Laboratory for *AI*-Powered Financial Technologies.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.artint.2025.104289.

Appendix A. – Proofs of Theorems, Lemmas and Corollaries

A1. Proof of Theorem 1 (MAKER rule)

$p(\theta|e_{li} \wedge e_{mj})$ in the Theorem is the conditional probability that is calculated in *SOPS* and measures the degree to which state θ is true given both evidence e_{li} and evidence e_{mj}. On the other hand, $P(\theta|e_{li} \wedge e_{mj})$ is the conditional probability that is calculated in *ERPS* and measures the degree to which state θ is supported given both evidence e_{li} and evidence e_{mj}. From Equation (13), Equation (17) and Equation (18), probability $p(\theta|e_{li} \wedge e_{mj})$ should be proportional to probability $P(\theta|e_{li} \wedge e_{mj})$, or

$$p(\theta|e_{li} \wedge e_{mj}) = k_1 P(\theta|e_{li} \wedge e_{mj}) \qquad \forall \theta \subseteq \Theta \tag{A1-1}$$

where k_1 is a positive constant that does not change for any $\theta \subseteq \Theta$.

$P(\theta|e_{li} \wedge e_{mj})$ is generated in the *ER* process of combining evidence e_{li} and evidence e_{mj} in *ERPS*, the first step of which is to calculate the joint probabilities that both evidence e_{li} and evidence e_{mj} support state θ in any legitimate way in *ERPS*. Let $(A|e_{li})$ stand for the event that evidence e_{li} supports state A and $(B|e_{mj})$ for the event that evidence e_{mj} supports state B. If $A \cap B = \theta$, joint event $(A|e_{li}) \cap (B|e_{mj})$ constitutes a piece of joint support for state θ from both e_{li} and e_{mj}. The total joint support for θ from e_{li} and e_{mj} is then the union of all exclusive joint events $(A|e_{li}) \cap (B|e_{mj})$ for any $A \cap B = \theta$ in *ERPS*. Probability $P(\theta|e_{li} \wedge e_{mj})$ is thus proportional to the joint probability of this union.

This union includes three distinctive parts in *ERPS*: joint support for $\theta : (A|e_{li}) \cap (B|e_{mj})$ for any $A, B \subseteq \Theta$ with $A \cap B = \theta$, e_{li}'s support for $\theta : (\theta|e_{li}) \cap (2^{\Theta}|e_{mj})$ that is gained from the residual support left over by e_{mj}, and e_{mj}'s support for $\theta : (2^{\Theta}|e_{li}) \cap (\theta|e_{mj})$ that is gained from the residual support left over by e_{li}, where $(2^{\Theta}|e_{li})$ and $(2^{\Theta}|e_{mj})$ are the residual support left over by e_{li} and e_{mj}, respectively. Note that the residual support of e_{li} is independent of e_{mj}'s support for θ for any $\theta \subseteq \Theta$, and vice versa. In other words, we have

$$P\big((\theta|e_{li})\cap(2^{\Theta}|e_{mj})\big)=P(\theta|e_{li})P(2^{\Theta}|e_{mj})\text{ and }P\big((2^{\Theta}|e_{li})\cap(\theta|e_{mj})\big)=P(2^{\Theta}|e_{li})P(\theta|e_{mj}) \tag{A1-2}$$

Let the joint probability of this union be denoted by m_θ. We then have

$$P(\theta|e_{li}\wedge e_{mj})=k_2m_\theta \qquad \forall\theta\subseteq\Theta\text{ and }\theta=2^{\Theta} \tag{A1-3}$$

where k_2 is a non-negative constant that does not change for any $\theta\subseteq\Theta$ and power set 2^{Θ}. k_2 needs to be determined from the requirement that the probabilities of joint support for all states and the power set be summed up to one.

Joint probability m_θ for any $\theta\subseteq\Theta$ is calculated as follows:

$$\begin{aligned}m_\theta&=P\left(\bigcup_{A\cap B=\theta}\big((A|e_{li})\cap(B|e_{mj})\big)\right)\\&=P\left(\left(\bigcup_{\substack{A,B\subseteq\Theta\\A\cap B=\theta}}\big((A|e_{li})\cap(B|e_{mj})\big)\right)\cup\big((\theta|e_{li})\cap(2^{\Theta}|e_{mj})\big)\cup\big((2^{\Theta}|e_{li})\cap(\theta|e_{mj})\big)\right)\\&=P\big((\theta|e_{li})\cap(2^{\Theta}|e_{mj})\big)+P\big((2^{\Theta}|e_{li})\cap(\theta|e_{mj})\big)+\sum_{\substack{A,B\subseteq\Theta\\A\cap B=\theta}}P\big((A|e_{li})\cap(B|e_{mj})\big)\end{aligned} \tag{A1-4}$$

From Equation (A1-2), Equation (A1-4), Equation (18), Equation (22), Equation (23), Equation (24), Equation (25) and Equation (26), we get

$$\begin{aligned}m_\theta&=P(\theta|e_{li})P(2^{\Theta}|e_{mj})+P(2^{\Theta}|e_{li})P(\theta|e_{mj})+\sum_{A,B\subseteq\Theta,A\cap B=\theta}P\big((A|e_{li})\cap(B|e_{mj})\big)\\&=\omega_{li}\omega_{mj}(1-r_{mj})m_{\theta li}+\omega_{li}\omega_{mj}(1-r_{li})m_{\theta mj}+\sum_{A,B\subseteq\Theta,A\cap B=\theta}p\big((s_{Alm}|e_{li})\cap(s_{Blm}|e_{mj})\big)\end{aligned} \tag{A1-5}$$

From Equation (8), Equation (18), Equation (19), Equation (20), Equation (22), Equation (23), Equation (47), Equation (49), Equation (51) and Equation (52), we can rewrite the probability of the last term in Equation (A1-5) as follows:

$$\begin{aligned}&p\big((s_{Alm}|e_{li})\cap(s_{Blm}|e_{mj})\big)=p\big(((A\cap e_{Alm})|e_{li})\cap((B\cap e_{Blm})|e_{mj})\big)\\&=p\big((A\cap B)\cap(e_{Alm}|e_{li})\cap(e_{Blm}|e_{mj})\big)=p\big(\theta\cap(e_{Alm}|e_{li})\cap(e_{Blm}|e_{mj})\big)\\&=p\big(\theta|(e_{Alm}|e_{li})\cap(e_{Blm}|e_{mj})\big)\,p\big((e_{Alm}|e_{li})\cap(e_{Blm}|e_{mj})\big)\\&=\frac{p\big(\theta|(e_{Alm}|e_{li})\cap(e_{Blm}|e_{mj})\big)}{p(A|e_{Al})p(B|e_{Bm})}\times\frac{\omega_{li,mj}p_{lm}\big((e_{Alm}|x_l=x_{li})\cap(e_{Blm}|x_m=x_{mj})\big)}{p_l(e_{Al}|x_l=x_{li})p_m(e_{Bm}|x_m=x_{mj})}\\&\times p(A|e_{Al})p_l(e_{Al}|x_l=x_{li})p(B|e_{Bm})p_m(e_{Bm}|x_m=x_{mj})\\&=\omega_{li,mj}\overline{w}_{Al,Bm}\overline{\alpha}_{Ali,Bmj}m_{Ali}m_{Bmj}\end{aligned} \tag{A1-6}$$

From Equation (A1-6) and Equation (54), Equation (A1-5) can be re-written as follows:

$$\begin{aligned}m_\theta&=\omega_{li}\omega_{mj}\big[(1-r_{mj})m_{\theta li}+(1-r_{li})m_{\theta mj}\big]+\omega_{li,mj}\sum_{\substack{A,B\subseteq\Theta\\A\cap B=\theta}}\overline{w}_{Al,Bm}\overline{\alpha}_{Ali,Bmj}m_{Ali}m_{Bmj}\\&=\omega_{li}\omega_{mj}\widehat{m}_{\theta(li,mj)}\qquad\forall\theta\subseteq\Theta\end{aligned} \tag{A1-7}$$

$$\widehat{m}_{\theta(li,mj)}=\big[(1-r_{mj})m_{\theta li}+(1-r_{li})m_{\theta mj}\big]+\overline{\omega}_{li,mj}\sum_{A,B\subseteq\Theta,A\cap B=\theta}\overline{w}_{Al,Bm}\overline{\alpha}_{Ali,Bmj}m_{Ali}m_{Bmj} \tag{A1-8}$$

Similarly, joint probability $m_{2^{\Theta}}$ for powerset 2^{Θ} is calculated from Equation (26) as follows by noting that the residual support of e_{li} is independent of that of e_{mj}, and vice versa,

$$m_{2^{\Theta}}=P\big((2^{\Theta}|e_{li})\cap(2^{\Theta}|e_{mj})\big)=P(2^{\Theta}|e_{li})P(2^{\Theta}|e_{mj})=\omega_{li}\omega_{mj}(1-r_{li})(1-r_{mj})=\omega_{li}\omega_{mj}\widehat{m}_{2^{\Theta}(li,mj)} \tag{A1-9}$$

$$\widehat{m}_{2^{\Theta}(li,mj)}=(1-r_{li})(1-r_{mj}) \tag{A1-10}$$

Since conditional probability $P(\theta|e_{li}\wedge e_{mj})$ is basic probability defined in Equation (17), according to Equation (A1-3), the following equation must hold:

$$\sum_{D\subseteq\Theta}P(D|e_{li}\wedge e_{mj})+P(2^{\Theta}|e_{li}\wedge e_{mj})=\sum_{D\subseteq\Theta}k_2m_D+k_2m_{2^{\Theta}}=1 \tag{A1-11}$$

Putting Equation (A1-7) and Equation (A1-9) into Equation (A1-11) leads to

$$k_2 = 1 \Bigg/ \left(\sum_{D \subseteq \Theta} m_D + m_{2^\Theta} \right) = 1 \Bigg/ \left(\omega_{li}\omega_{mj} \left(\sum_{D \subseteq \Theta} \widehat{m}_{D(li,mj)} + \widehat{m}_{2^\Theta(li,mj)} \right) \right) \tag{A1-12}$$

From Equation (A1-3), we can use the constant k_2 generated by Equation (A1-12) to fulfil the second step of the *ER* process of calculating $P(\theta | e_{li} \wedge e_{mj})$ as conditional probability by

$$\begin{aligned} P(\theta | e_{li} \wedge e_{mj}) &= k_2 m_\theta = m_\theta \Bigg/ \left(\sum_{D \subseteq \Theta} m_D + m_{2^\Theta} \right) \\ &= \omega_{li}\omega_{mj}\widehat{m}_{\theta(li,mj)} \Bigg/ \left(\sum_{D \subseteq \Theta} \omega_{li}\omega_{mj}\widehat{m}_{D(li,mj)} + \omega_{li}\omega_{mj}\widehat{m}_{2^\Theta(li,mj)} \right) \\ &= \widehat{m}_{\theta(li,mj)} \Bigg/ \left(\sum_{D \subseteq \Theta} \widehat{m}_{D(li,mj)} + \widehat{m}_{2^\Theta(li,mj)} \right) \forall \theta \subseteq \Theta \end{aligned} \tag{A1-13}$$

$$\begin{aligned} P(2^\Theta | e_{li} \wedge e_{mj}) &= k_2 m_{2^\Theta} = m_{2^\Theta} \Bigg/ \left(\sum_{D \subseteq \Theta} m_D + m_{2^\Theta} \right) \\ &= \omega_{li}\omega_{mj}\widehat{m}_{2^\Theta(li,mj)} \Bigg/ \left(\sum_{D \subseteq \Theta} \omega_{li}\omega_{mj}\widehat{m}_{D(li,mj)} + \omega_{li}\omega_{mj}\widehat{m}_{2^\Theta(li,mj)} \right) \\ &= \widehat{m}_{2^\Theta(li,mj)} \Bigg/ \left(\sum_{D \subseteq \Theta} \widehat{m}_{D(li,mj)} + \widehat{m}_{2^\Theta(li,mj)} \right) \end{aligned} \tag{A1-14}$$

Since $p(\theta | e_{li} \wedge e_{mj})$ is basic probability defined in Equation (1), according to Equation (A1-1) and Equation (A1-13), the following equation must hold:

$$\sum_{C \subseteq \Theta} p(C | e_{li} \wedge e_{mj}) = \sum_{C \subseteq \Theta} k_1 P(C | e_{li} \wedge e_{mj}) = k_1 \sum_{C \subseteq \Theta} \left(\widehat{m}_{C(li,mj)} \Bigg/ \left(\sum_{D \subseteq \Theta} \widehat{m}_{D(li,mj)} + \widehat{m}_{2^\Theta(li,mj)} \right) \right) = 1 \tag{A1-15}$$

leading to

$$k_1 = \frac{\sum_{D \subseteq \Theta} \widehat{m}_{D(li,mj)} + \widehat{m}_{2^\Theta(li,mj)}}{\sum_{C \subseteq \Theta} \widehat{m}_{C(li,mj)}} \tag{A1-16}$$

From Equation (A1-1), we then have

$$\begin{aligned} p(\theta | e_{li} \wedge e_{mj}) &= k_1 \; P(\theta | e_{li} \wedge e_{mj}) \\ &= \frac{\sum_{D \subseteq \Theta} \widehat{m}_{D(li,mj)} + \widehat{m}_{2^\Theta(li,mj)}}{\sum_{C \subseteq \Theta} \widehat{m}_{C(li,mj)}} \left(\widehat{m}_{\theta(li,mj)} \Bigg/ \left(\sum_{D \subseteq \Theta} \widehat{m}_{D(li,mj)} + \widehat{m}_{2^\Theta(li,mj)} \right) \right) \\ &= \widehat{m}_{\theta(li,mj)} \Bigg/ \sum_{C \subseteq \Theta} \widehat{m}_{C(li,mj)} \forall \theta \subseteq \Theta \end{aligned} \tag{A1-17}$$

From Equation (A1-10) and Equation (A1-14), the untrustworthiness of the results is given by

$$\begin{aligned} \widetilde{m}_{2^\Theta(li,mj)} &= P(2^\Theta | e_{li} \wedge e_{mj}) = \widehat{m}_{2^\Theta(li,mj)} \Bigg/ \left(\sum_{D \subseteq \Theta} \widehat{m}_{D(li,mj)} + \widehat{m}_{2^\Theta(li,mj)} \right) \\ &= (1 - r_{li})(1 - r_{mj}) \Bigg/ \left(\sum_{D \subseteq \Theta} \widehat{m}_{D(li,mj)} + (1 - r_{li})(1 - r_{mj}) \right) \end{aligned} \tag{A1-18}$$

Finally, if $\theta = \emptyset$, there must be $p(\emptyset | e_{li} \wedge e_{mj}) = 0$ by Definition 1.

A2. Proof of Corollary 1.1 (Exclusive MAKER rule)

From Definition 13 or Definition 14 and Equation (23), by putting Equation (55) or Equation (56) into Equation (A1-5), Equation (A1-8) reduces to

$$\widehat{m}_{\theta(li,mj)} = \left[(1 - r_{mj}) m_{\theta li} + (1 - r_{li}) m_{\theta mj} \right] = \left[(1 - r_{mj}) w_{\theta l} p_{\theta li} + (1 - r_{li}) w_{\theta m} p_{\theta mj} \right] \tag{A2-1}$$

Putting the above equation into Equation (59) leads to

$$p(\theta|e_{li} \wedge e_{mj}) = \frac{(1-r_{mj})w_{\theta l}p_{\theta li} + (1-r_{li})w_{\theta m}p_{\theta mj}}{(1-r_{mj})\sum_{C\subseteq\Theta} m_{Cli} + (1-r_{li})\sum_{C\subseteq\Theta} m_{Cmj}} \qquad \forall\theta \subseteq \Theta$$

A3. Proof of Corollary 1.2 (Exclusive ER rule)

In Corollary 1.1, if the weights of variables remain constant for all state, or $w_{\theta l} = w_l$ and $w_{\theta m} = w_m$ for any $\theta \subseteq \Theta$, from Equation (8) and Equation (23), we have

$$\sum_{C\subseteq\Theta} m_{Cli} = \sum_{C\subseteq\Theta} w_{Cl}p_{Cli} = w_l \sum_{C\subseteq\Theta} p_{Cli} = w_l \tag{A3-1}$$

$$\sum_{C\subseteq\Theta} m_{Cmj} = \sum_{C\subseteq\Theta} w_{Cm}p_{Cmj} = w_m \sum_{C\subseteq\Theta} p_{Cmj} = w_m \tag{A3-2}$$

Putting Equation (A3-1) and Equation (A3-2) into Equation (62) leads to

$$p(\theta|e_{li} \wedge e_{mj}) = \frac{(1-r_{mj})w_l}{(1-r_{mj})w_l + (1-r_{li})w_m}p_{\theta li} + \frac{(1-r_{li})w_m}{(1-r_{mj})w_l + (1-r_{li})w_m}p_{\theta mj} \qquad \forall\theta \subseteq \Theta$$

A4. Proof of Corollary 1.3 (Differential MAKER rule)

Since evidence e_{li} and evidence e_{mj} are exclusive of each other and each fully reliable, we have $r_{li} = r_{mj} = 1$ but cannot put them into Equation (62) to calculate the combined likelihood because both the numerator and denominator of Equation (62) are zero so it is not defined in this case. However, we conjecture that r_{li} (r_{mj}) are the functions of parameter t in a small area around $r_{li} = 1$ ($r_{mj} = 1$), and define r_{li} and r_{mj} as some kind of differentiable functions of t:

$$r_{li} = r_{li}(t) \text{ and } r_{mj} = r_{mj}(t) \text{ with } r_{li}(t=0) = 1 \text{ and } r_{mj}(t=0) = 1 \tag{A4-1}$$

Putting $r_{li}(t)$ and $r_{mj}(t)$ into Equation (60), we get

$$\widehat{m}_{C(li,mj)}(t) = (1-r_{mj}(t))m_{Cli} + (1-r_{li}(t))m_{Cmj} \tag{A4-2}$$

We can then calculate $p(\theta|e_{li} \wedge e_{mj})$ by finding the following limit (Taylor, 1952)

$$p(\theta|e_{li} \wedge e_{mj}) = \lim_{t\to 0}\frac{\widehat{m}_{\theta(li,mj)}(t)}{\sum_{C\subseteq\Theta}\widehat{m}_{C(li,mj)}(t)} = \lim_{t\to 0}\frac{\widehat{m}_{\theta(li,mj)}(t)/t}{\sum_{C\subseteq\Theta}\widehat{m}_{C(li,mj)}(t)/t} = \frac{d(\widehat{m}_{\theta(li,mj)}(t))/dt}{\sum_{C\subseteq\Theta} d(\widehat{m}_{C(li,mj)}(t))/dt}\bigg|_{t=0} \tag{A4-3}$$

Derivative $d(\widehat{m}_{C(li,mj)}(t))/dt$ is calculated as follows:

$$d(\widehat{m}_{C(li,mj)}(t))\Big/dt = d((1-r_{mj}(t))m_{Cli} + (1-r_{li}(t))m_{Cmj})\Big/dt = -r'_{mj}(t)m_{Cli} - r'_{li}(t)m_{Cmj} \tag{A4-4}$$

Putting Equation (A4-4) into Equation (A4-3) leads to

$$p(\theta|e_{li} \wedge e_{mj}) = \frac{-r'_{mj}(t)m_{\theta li} - r'_{li}(t)m_{\theta mj}}{\sum_{C\subseteq\Theta}\left(-r'_{mj}(t)m_{Cli} - r'_{li}(t)m_{Cmj}\right)}\Bigg|_{t=0} = \frac{r'_{mj}(0)m_{\theta li} + r'_{li}(0)m_{\theta mj}}{\sum_{C\subseteq\Theta}\left(r'_{mj}(0)m_{Cli} + r'_{li}(0)m_{Cmj}\right)} \tag{A4-5}$$

where $r'_{li}(0)$ is the derivative of the reliability of e_{li} at $r_{li}(t=0) = 1$. Let $r'_{li} = r'_{li}(0)$ and $r'_{mj} = r'_{mj}(0)$. Since r'_{li} and r'_{mj} are constant for any $\theta \subseteq \Theta$, from Equation (23) we then have

$$p(\theta|e_{li} \wedge e_{mj}) = \frac{r'_{mj}w_{\theta l}p_{\theta li} + r'_{li}w_{\theta m}p_{\theta mj}}{r'_{mj}\sum_{C\subseteq\Theta} m_{Cli} + r'_{li}\sum_{C\subseteq\Theta} m_{Cmj}} \tag{A4-6}$$

For $\theta = \emptyset$, there must be $p(\theta|e_{li} \wedge e_{mj}) = 0$

A5. Proof of Corollary 1.4 (Independent MAKER rule)

If evidence e_{li} and evidence e_{mj} are mutually independent, from Equation (20), Equation (23) and Equation (57), Equation (A1-6) of Appendix A1 becomes

$$p((s_{Alm}|e_{li}) \cap (s_{Blm}|e_{mj})) = p(s_{Al}|e_{li})p(s_{Bm}|e_{mj}) = p(A \cap e_{Al}|e_{li})p(B \cap e_{Bm}|e_{mj})$$

$$= p(A|e_{Al})p(e_{Al}|e_{li})p(B|e_{Bm})p(e_{Bm}|e_{mj})$$

$$= p(A|e_{Al})\omega_{li}p_l(e_{Al}|x_l = x_{li})p(B|e_{Bm})\omega_{mj}p_m(e_{Bm}|x_m = x_{mj})$$

$$= \omega_{li}\omega_{mj}m_{Ali}m_{Bmj}$$

Putting the above equation into Equation (A1-5) of Appendix A1 and following the same proof process from Equation (A1-6) in Appendix A1 lead to the conclusion of *Corollary 1.4*.

A6. *Proof of Corollary 1.5 (ER rule)*

If $w_{\theta l} = w_l$ and $w_{\theta m} = w_m$ for any $\theta \subseteq \Theta$, Equation (67) reduces to Equation (40) of reference [71], which confirms that the independent *MAKER* rule reduces to the *ER* rule in this case.

A7. *Proof of Corollary 1.6 (Augmented Dempster's rule)*

For $\theta = \varnothing$, from Equation (1) there must be $p(\theta|e_{li} \wedge e_{mj}) = 0$.

For any $\sigma > \tau$, the orthogonal sum of joint probabilities is regarded to be appropriate and meaningful in that Dempster's rule is adequately defined in this case and it is legitimate to apply it. In Corollary 1.5, if each piece of evidence is fully reliable, there are $r_{li} = r_{mj} = 1$. By putting them into Equation (67) that is in turn put into Equation (59) and noting Equation (23) and $w_{\theta l} = w_l$ and $w_{\theta m} = w_m$ for any $\theta \subseteq \Theta$, Theorem 1 reduces to

$$p(\theta|e_{li} \wedge e_{mj}) = \frac{\sum_{A\cap B=\theta} m_{Ali}m_{Bmj}}{\sum_{C\subseteq\Theta}\sum_{A\cap B=C} m_{Ali}m_{Bmj}} = \frac{\sum_{A\cap B=\theta} w_{Al}p_{Ali}w_{Bm}p_{Bmj}}{\sum_{C\subseteq\Theta}\sum_{A\cap B=C} w_{Al}p_{Ali}w_{Bm}p_{Bmj}}$$
$$= \frac{\sum_{A\cap B=\theta} w_l p_{Ali}w_m p_{Bmj}}{\sum_{C\subseteq\Theta}\sum_{A\cap B=C} w_l p_{Ali}w_m p_{Bmj}} = \frac{\sum_{A\cap B=\theta} p_{Ali}p_{Bmj}}{\sum_{C\subseteq\Theta}\sum_{A\cap B=C} p_{Ali}p_{Bmj}} \quad \theta \subseteq \Theta, \sigma > \tau \tag{A7-1}$$

From Definition 1 and Definition 4, there is $\sum_{C\subseteq\Theta}\sum_{A\cap B=C} p_{Ali}p_{Bmj} + \sum_{A\cap B=\varnothing} p_{Ali}p_{Bmj} = 1$. We therefore have

$$\sum_{C\subseteq\Theta}\sum_{A\cap B=C} p_{Ali}p_{Bmj} = 1 - \sum_{A\cap B=\varnothing} p_{Ali}p_{Bmj} \tag{A7-2}$$

Putting Equation (A7-2) into Equation (A7-1) leads to

$$p(\theta|e_{li} \wedge e_{mj}) = \frac{\sum_{A\cap B=\theta} p_{Ali}p_{Bmj}}{1 - \sum_{A\cap B=\varnothing} p_{Ali}p_{Bmj}} \quad \theta \subseteq \Theta, \sigma > \tau \tag{A7-3}$$

For $\tau \geq \sigma > 0$, the orthogonal sum of joint probability is deemed to be too small to mean likely. It is therefore not appropriate to apply Dempster's rule given by Equation (A7-1) or Equation (A7-3) to calculate $p(\theta|e_{li} \wedge e_{mj})$ as in this case Equation (A7-1) is no longer defined in a meaningful way. As such, we attempt to perturb the 100% reliability for evidence in order to identify how evidence should be combined in a small neighbourhood of the 100% reliability. Such perturbation is to reduce the reliability of evidence by a very small amount, e.g. δ_{li} and δ_{mj} for e_{li} and e_{mj}, or $1 \geq r_{li} \geq 1 - \delta_{li}$ and $1 \geq r_{mj} \geq 1 - \delta_{mj}$ with δ_{li}, $\delta_{mj} \ll \tau$, so that individual support for states can be counted for in the very small neighbourhood. To perturb evidence reliability, a reliability perturbation function $r_{li}(\sigma)$ for evidence e_{li} needs to be constructed for each piece of evidence, which should satisfy certain conditions.

$r_{li}(\sigma)$ should meet at least four conditions: i) differentiable, ii) $r_{li}(\tau) = 1$, iii) $r_{li}(0) = 1$ and iv) $1 \geq r_{li}(\sigma) \geq 1 - \delta_{li}$ for any $\tau \geq \sigma > 0$. Condition i) is required to ensure that Dempster's rule can be defined when evidence is in complete conflict as analysed below; Conditions ii) and iii) are needed to ensure the continuity of reliability at $\sigma = \tau$ and $\sigma = 0$. Condition iv) is required to ensure that perturbation is controlled to be no more than δ_{li} so that any reduction of $r_{li}(\sigma)$ from the 100% reliability can be regarded negligible.

A general quadratic reliability perturbation function $r_{li}(\sigma) = a + b\sigma + c\sigma^2$ is constructed to meet the conditions precisely. The fourth condition requires that the minimum reliability for evidence e_{li} be equal to $1 - \delta_{li}$, which occurs at $\sigma = -b/(2c)$. This fourth condition plus Conditions ii) and iii) leads to the following three equations to determine parameters *a, b* and *c*:

$$r_{li}(\tau) = a + b\tau + c\tau^2 = 1 \tag{A7-4}$$

$$r_{li}(0) = a = 1 \tag{A7-5}$$

$$r_{li}(-b/(2c)) = a + b(-b/(2c)) + c(-b/(2c))^2 = 1 - \delta_{li} \tag{A7-6}$$

Solving the above equations leads to $a = 1$, $b = -4\delta_{li}/\tau$ and $c = 4\delta_{li}/\tau^2$, so we get

$$r_{li}(\sigma) = 1 - (4\delta_{li}/\tau)\sigma + (4\delta_{li}/\tau^2)\sigma^2 \tag{A7-7}$$

Similarly, we can construct a quadratic reliability perturbation function $r_{mj}(\sigma)$ for evidence e_{mj}

$$r_{mj}(\sigma) = 1 - (4\delta_{mj}/\tau)\sigma + (4\delta_{mj}/\tau^2)\sigma^2 \tag{A7-8}$$

If more conditions need to be met, higher order polynomial reliability perturbation functions can be constructed in a similar way. Equation (67) then becomes

$$\begin{aligned}\widehat{m}_{\theta(li,mj)} &= \left[(1-r_{mj}(\sigma))w_{\theta l}p_{\theta li} + (1-r_{li}(\sigma))w_{\theta m}p_{\theta mj}\right] + \sum_{\substack{A,B\subseteq\Theta\\A\cap B=\theta}} w_{Al}p_{Ali}w_{Bm}p_{Bmj}\\ &= \left[(1-r_{mj}(\sigma))w_l p_{\theta li} + (1-r_{li}(\sigma))w_m p_{\theta mj}\right] + w_l w_m \sum_{\substack{A,B\subseteq\Theta\\A\cap B=\theta}} p_{Ali}p_{Bmj}\\ &= w_l w_m \overline{m}_\theta \qquad \forall\theta\subseteq\Theta\end{aligned} \tag{A7-9}$$

$$\overline{m}_\theta = \left[\frac{1-r_{mj}(\sigma)}{w_m}p_{\theta li} + \frac{1-r_{li}(\sigma)}{w_l}p_{\theta mj}\right] + \sum_{\substack{A,B\subseteq\Theta\\A\cap B=\theta}} p_{Ali}p_{Bmj} \tag{A7-10}$$

Putting Equation (A7-9) into Equation (59) leads to

$$p(\theta|e_{li}\wedge e_{mj}) = \frac{\widehat{m}_{\theta(li,mj)}}{\sum_{C\subseteq\Theta}\widehat{m}_{C(li,mj)}} = \frac{w_l w_m \overline{m}_\theta}{\sum_{C\subseteq\Theta}(w_l w_m \overline{m}_C)} = \frac{\overline{m}_\theta}{\sum_{C\subseteq\Theta}\overline{m}_C} \tag{A7-11}$$

For $\sigma = 0$, we have the same situation as analysed in Corollary 1.3. From Equation (A7-7), we get $r'_{li}(\sigma) = -(4\delta_{li}/\tau) + 2(4\delta_{li}/\tau^2)\sigma$, so $r'_{li}(0) = -(4\delta_{li}/\tau)$. Similar, we get $r'_{mj}(0) = -(4\delta_{mj}/\tau)$. Putting them into Equation (66) leads to

$$\begin{aligned}p(\theta|e_{li}\wedge e_{mj}) &= \frac{-(4\delta_{mj}/\tau)w_{\theta l}p_{\theta li} - (4\delta_{li}/\tau)w_{\theta m}p_{\theta mj}}{-(4\delta_{mj}/\tau)\sum_{C\subseteq\Theta}w_{Cl}p_{Cli} - (4\delta_{li}/\tau)\sum_{D\subseteq\Theta}w_{Dm}p_{Dmj}}\\ &= \frac{\delta_{mj}w_l p_{\theta li} + \delta_{li}w_m p_{\theta mj}}{\delta_{mj}w_l\sum_{C\subseteq\Theta}p_{Cli} + \delta_{li}w_m\sum_{D\subseteq\Theta}p_{Dmj}}\end{aligned} \tag{A7-12}$$

From Definition (3), $\sum_{C\subseteq\Theta}p_{Cli} = 1$ and $\sum_{D\subseteq\Theta}p_{Dmj} = 1$. Equation (A7-12) finally becomes

$$p(\theta|e_{li}\wedge e_{mj}) = \frac{\delta_{mj}w_l}{\delta_{mj}w_l + \delta_{li}w_m}p_{\theta li} + \frac{\delta_{li}w_m}{\delta_{mj}w_l + \delta_{li}w_m}p_{\theta mj}$$

A8. *Proof of Lemma 2.1 (Recursive MAKER algorithm)*

Let $i_1 = i-1$ and $e_{1\overline{\wedge}i_1} = (e_1\wedge\cdots\wedge e_{i-1})$ be the conjunction of the first *i*-1 pieces of evidence, and $e_{\theta(1\overline{\wedge}i_1)}$ be the evidential element that evidence $e_{1\overline{\wedge}i_1}$ points to state θ.

For $\theta = \emptyset$, from Definition (6), there must be $P(\emptyset|e_{1\overline{\wedge}l}) = 0$ for any l.

For $\theta \neq \emptyset$, at l=2, given the above notations, using Equation (A1-8), Equation (A1-10), Equation (A1-13) and Equation (A1-14) in the proof of Theorem 1 to combine two pieces of evidence e_1 and e_2 leads to,

$$\widetilde{m}_{\theta(1\overline{\wedge}2)} = P(\theta|e_{1\overline{\wedge}2}) = \frac{\widehat{m}_{\theta(1\overline{\wedge}2)}}{\sum_{C\subseteq\Theta}\widehat{m}_{C(1\overline{\wedge}2)} + \widehat{m}_{2^\Theta(1\overline{\wedge}2)}} \qquad \forall\theta\neq\emptyset \tag{A8-1}$$

with $\widehat{m}_{\theta(1\overline{\wedge}2)}$ for any $\theta\subseteq\Theta$ and $\widehat{m}_{2^\Theta(1\overline{\wedge}2)}$ given as follows

$$\widehat{m}_{\theta(1\overline{\wedge}2)} = [\widetilde{m}_{\theta(1\overline{\wedge}1)}(1-r_2) + \widetilde{m}_{2^\Theta(1\overline{\wedge}1)}m_{\theta 2}] + \overline{\omega}_{1,2}\sum_{\substack{A,B\subseteq\Theta\\A\cap B=\theta}}\overline{w}_{A1,B2}\overline{\alpha}_{A1,B2}\widetilde{m}_{A(1\overline{\wedge}1)}m_{B2} \quad \forall\theta\subseteq\Theta \tag{A8-2}$$

$$\widehat{m}_{2^\Theta(1\overline{\wedge}2)} = (1-r_2)\widetilde{m}_{2^\Theta(1\overline{\wedge}1)} \tag{A8-3}$$

where $\widetilde{m}_{\theta(1\overline{\wedge}1)} = \omega_1 m_{\theta 1}$ and $\widetilde{m}_{2^\Theta(1\overline{\wedge}1)} = \omega_1(1-r_1)$ as given by Equation (22) and Equation (26), respectively. Since there is ω_1 in every term in Equation (A8-2) and (A8-3), ω_1 will be cancelled out in Equation (A8-1). To simplify calculation, we set

$$\widetilde{m}_{\theta(1\overline{\wedge}1)} = m_{\theta 1} \text{ and } \widetilde{m}_{2^\Theta(1\overline{\wedge}1)} = (1-r_1) \tag{A8-4}$$

without changing the result of Equation (A8-1). In Equation (A8-2), $\overline{\omega}_{1,2} = \omega_{1,2}/(\omega_1\omega_2)$, $\overline{w}_{A1,B2} = w_{\theta 12}/(w_{A1}w_{B2})$, $\overline{\alpha}_{A1,B2} = p_{A1,B2}/(p_{A1}p_{B2})$ with $A\cap B = \theta$. So, Equation (70) to Equation (72) hold at l=2.

Assume that Equation (70) to Equation (72) hold at $l=i_1$. Let $s_{\theta(1\overline{\wedge}i_1)} = (\theta \cap e_{\theta(1\overline{\wedge}i_1)})$ and $(\theta|e_{1\overline{\wedge}i_1})$ stand for evidence $e_{1\overline{\wedge}i_1}$ supporting state θ in *SOPS* and *ERPS*, respectively, with $\widetilde{m}_{\theta(1\overline{\wedge}i_1)} = P(\theta|e_{1\overline{\wedge}i_1}) = p(s_{\theta(1\overline{\wedge}i_1)})$ as given in Equation (17) and Equation (22). From the above assumption and notations, we then have

$$\widetilde{m}_{2^\Theta(1\overline{\wedge}i_1)} = P(2^\Theta|e_{1\overline{\wedge}i_1}) = \frac{\widehat{m}_{2^\Theta(1\overline{\wedge}i_1)}}{\sum_{C\subseteq\Theta}\widehat{m}_{C(1\overline{\wedge}i_1)} + \widehat{m}_{2^\Theta(1\overline{\wedge}i_1)}} \tag{A8-5}$$

$$\widetilde{m}_{\theta(1\overline{\wedge}i_1)} = P(\theta|e_{1\overline{\wedge}i_1}) = p(s_{\theta(1\overline{\wedge}i_1)}) = p(\theta\cap e_{\theta(1\overline{\wedge}i_1)}) = p(\theta|e_{\theta(1\overline{\wedge}i_1)})p(e_{\theta(1\overline{\wedge}i_1)}) = \frac{\widehat{m}_{\theta(1\overline{\wedge}i_1)}}{\sum_{C\subseteq\Theta}\widehat{m}_{C(1\overline{\wedge}i_1)} + \widehat{m}_{2^\Theta(1\overline{\wedge}i_1)}} \quad \forall\theta\subseteq\Theta \tag{A8-6}$$

At $l=i$, we need to generate basic probability $\widetilde{m}_{\theta(1\overline{\wedge}i)} = P(\theta|e_{1\overline{\wedge}i})$ for any $\theta\neq\emptyset$, which should be proportional to the joint probability of the union of the exclusive events that both evidence $e_{1\overline{\wedge}i_1}$ supports state A and evidence e_i supports state B, that is, $(A|e_{1\overline{\wedge}i_1})\cap(B|e_i)$ for any $A\cap B = \theta$. In other words, the following equation should hold:

$$\widetilde{m}_{\theta(1\overline{\wedge}i)} = P(\theta|e_{1\overline{\wedge}i}) = k_3 P\left(\bigcup_{A\cap B=\theta}((A|e_{1\overline{\wedge}i_1})\cap(B|e_i))\right) \tag{A8-7}$$

where k_3 is a non-negative constant that does not change for any $\theta\subseteq\Theta$ and power set 2^Θ.

Note that the residual support $(2^\Theta|e_{1\overline{\wedge}i_1})$ of evidence $e_{1\overline{\wedge}i_1}$ is formed independently of evidence e_i, and the residual support $(2^\Theta|e_i)$ of evidence e_i is formed independently of any evidence in $e_{1\overline{\wedge}i_1}$. For $\theta = 2^\Theta$, joint probability $P\left(\bigcup_{A\cap B=\theta}((A|e_{1\overline{\wedge}i_1})\cap(B|e_i))\right)$ is therefore calculated by Equation (26) and Equation (A8-5) as follows:

$$\begin{aligned} &P\left(\bigcup_{A\cap B=2^\Theta}((A|e_{1\overline{\wedge}i_1})\cap(B|e_i))\right) = P((2^\Theta|e_{1\overline{\wedge}i_1})\cap(2^\Theta|e_i)) \\ &= P(2^\Theta|e_{1\overline{\wedge}i_1})P(2^\Theta|e_i) = \widetilde{m}_{2^\Theta(1\overline{\wedge}i_1)}\omega_i(1-r_i) = \omega_i\widehat{m}_{2^\Theta(1\overline{\wedge}i)} \end{aligned} \tag{A8-8}$$

$$\widehat{m}_{2^\Theta(1\overline{\wedge}i)} = \widetilde{m}_{2^\Theta(1\overline{\wedge}i_1)}(1-r_i) \tag{A8-9}$$

For any $\theta\subseteq\Theta$, joint probability $P\left(\bigcup_{A\cap B=\theta,\theta\subseteq\Theta}((A|e_{1\overline{\wedge}i_1})\cap(B|e_i))\right)$ is calculated by Equation (22), Equation (26), Equation (A8-5) and Equation (A8-6) as follows:

$$\begin{aligned} &P\left(\bigcup_{\substack{A\cap B=\theta\\ \theta\subseteq\Theta}}((A|e_{1\overline{\wedge}i_1})\cap(B|e_i))\right) = P\left(\bigcup_{\substack{A\cap B=\theta\\ A,\ B\subseteq\Theta}}((A|e_{1\overline{\wedge}i_1})\cap(B|e_i))\cup((\theta|e_{1\overline{\wedge}i_1})\cap(2^\Theta|e_i))\cup((2^\Theta|e_{1\overline{\wedge}i_1})\cap(\theta|e_i))\right) \\ &= \sum_{\substack{A\cap B=\theta\\ A,\ B\subseteq\Theta}} P((A|e_{1\overline{\wedge}i_1})\cap(B|e_i)) + P((\theta|e_{1\overline{\wedge}i_1})\cap(2^\Theta|e_i)) + P((2^\Theta|e_{1\overline{\wedge}i_1})\cap(\theta|e_i)) = P(\theta|e_{1\overline{\wedge}i_1})P(2^\Theta|e_i) + P(2^\Theta|e_{1\overline{\wedge}i_1})P(\theta|e_i) \\ &\quad + \sum_{\substack{A\cap B=\theta\\ A,\ B\subseteq\Theta}} P((A|e_{1\overline{\wedge}i_1})\cap(B|e_i)) = \widetilde{m}_{\theta(1\overline{\wedge}i_1)}\omega_i(1-r_i) + \widetilde{m}_{2^\Theta(1\overline{\wedge}i_1)}\omega_i m_{\theta i} + \sum_{\substack{A\cap B=\theta\\ A,\ B\subseteq\Theta}} P((A|e_{1\overline{\wedge}i_1})\cap(B|e_i)) \end{aligned} \tag{A8-10}$$

The probability in the summation of the above equation can be rewritten as follows:

$$\begin{aligned}
&P((A|e_{1\overline{\wedge}i_1})\cap(B|e_i))=p(s_{A(1\overline{\wedge}i_1)}\cap s_{Bi})=p((A\cap e_{A(1\overline{\wedge}i_1)})\cap(B\cap e_{Bi}))\\
&=p((A\cap B)\cap(e_{A(1\overline{\wedge}i_1)}\cap e_{Bi}))=p(\theta\cap(e_{A(1\overline{\wedge}i_1)}\cap e_{Bi}))\\
&=p(\theta|e_{A(1\overline{\wedge}i_1)}\cap e_{Bi})p(e_{A(1\overline{\wedge}i_1)}\cap e_{Bi})\\
&=\frac{p(\theta|e_{A(1\overline{\wedge}i_1)}\cap e_{Bi})}{p(A|e_{A(1\overline{\wedge}i_1)})p(B|e_{Bi})}\times\frac{p(e_{A(1\overline{\wedge}i_1)}\cap e_{Bi})}{p(e_{A(1\overline{\wedge}i_1)})p_i(e_{Bi})}\\
&\times p(A|e_{A(1\overline{\wedge}i_1)})p(B|e_{Bi})\times p(e_{A(1\overline{\wedge}i_1)})p_i(e_{Bi})\\
&=\overline{w}_{A(1\overline{\wedge}i_1),Bi}\times\frac{\omega_{1\overline{\wedge}i}p_{1\overline{\wedge}i}(e_{A(1\overline{\wedge}i_1)}\cap e_{Bi})}{\omega_{1\overline{\wedge}i_1}p_{1\overline{\wedge}i_1}(e_{A(1\overline{\wedge}i_1)})p_i(e_{Bi})}\\
&\times p(A|e_{A(1\overline{\wedge}i_1)})p(e_{A(1\overline{\wedge}i_1)})\times p(B|e_{Bi})p_i(e_{Bi})\\
&=\frac{\omega_{1\overline{\wedge}i}}{\omega_{1\overline{\wedge}i_1}}\overline{w}_{A(1\overline{\wedge}i_1),Bi}\overline{\alpha}_{A(1\overline{\wedge}i_1),Bi}\widetilde{m}_{A(1\overline{\wedge}i_1)}m_{Bi}
\end{aligned} \tag{A8-11}$$

where

$$p(e_{A(1\overline{\wedge}i_1)}\cap e_{Bi})=\omega_{1\overline{\wedge}i}p_{1\overline{\wedge}i}(e_{A(1\overline{\wedge}i_1)}\cap e_{Bi}) \tag{A8-12}$$

$$p(e_{A(1\overline{\wedge}i_1)})=\omega_{1\overline{\wedge}i_1}p_{1\overline{\wedge}i_1}(e_{A(1\overline{\wedge}i_1)}) \tag{A8-13}$$

$$\overline{\alpha}_{A(1\overline{\wedge}i_1),Bi}=\frac{p_{1\overline{\wedge}i}(e_{A(1\overline{\wedge}i_1)}\cap e_{Bi})}{p_{1\overline{\wedge}i_1}(e_{A(1\overline{\wedge}i_1)})p_i(e_{Bi})} \tag{A8-14}$$

$$\overline{w}_{A(1\overline{\wedge}i_1),Bi}=p(\theta|e_{A(1\overline{\wedge}i_1)}\cap e_{Bi})/(p(A|e_{A(1\overline{\wedge}i_1)})p(B|e_{Bi})) \tag{A8-15}$$

$$\widetilde{m}_{A(1\overline{\wedge}i_1)}=\ p(A|e_{A(1\overline{\wedge}i_1)})p(e_{A(1\overline{\wedge}i_1)}) \tag{A8-16}$$

$$m_{Bi}=p(B|e_{Bi})p_i(e_{Bi}) \tag{A8-17}$$

with $p_{1\overline{\wedge}i}$, $p_{1\overline{\wedge}i_1}$ and p_i being the basic probability functions generated from the data sources where evidence $e_{1\overline{\wedge}i}$, evidence $e_{1\overline{\wedge}i_1}$ and evidence e_i are acquired, respectively.

Putting Equation (A8-11) into Equation (A8-10) leads to

$$P\left(\bigcup_{\substack{A\cap B=\theta\\ \theta\subseteq\Theta}}((A|e_{1\overline{\wedge}i_1})\cap(B|e_i))\right)=\omega_i[\widetilde{m}_{\theta(1\overline{\wedge}i_1)}(1-r_i)+\widetilde{m}_{2^\Theta(1\overline{\wedge}i_1)}m_{\theta i}]+\frac{\omega_{1\overline{\wedge}i}}{\omega_{1\overline{\wedge}i_1}}\sum_{\substack{A\cap B=\theta\\ A,\ B\subseteq\Theta}}\overline{w}_{A(1\overline{\wedge}i_1),Bi}\overline{\alpha}_{A(1\overline{\wedge}i_1),Bi}\widetilde{m}_{A(1\overline{\wedge}i_1)}m_{Bi}=\omega_i\widehat{m}_{\theta(1\overline{\wedge}i)},\forall\theta\subseteq\Theta \tag{A8-18}$$

$$\widehat{m}_{\theta(1\overline{\wedge}i)}=[\widetilde{m}_{\theta(1\overline{\wedge}i_1)}(1-r_i)+\widetilde{m}_{2^\Theta(1\overline{\wedge}i_1)}m_{\theta i}]+\overline{\omega}_{1\overline{\wedge}i_1,i}\sum_{\substack{A\cap B=\theta\\ A,\ B\subseteq\Theta}}\overline{w}_{A(1\overline{\wedge}i_1),Bi}\overline{\alpha}_{A(1\overline{\wedge}i_1),Bi}\widetilde{m}_{A(1\overline{\wedge}i_1)}m_{Bi} \tag{A8-19}$$

$$\overline{\omega}_{1\overline{\wedge}i_1,i}=\frac{\omega_{1\overline{\wedge}i}}{\omega_{1\overline{\wedge}i_1}\omega_i} \tag{A8-20}$$

Putting Equation (A8-8) and Equation (A8-18) into Equation (A8-7) and satisfying the condition $\sum\limits_{C\subseteq\Theta}P(C|e_{1\overline{\wedge}i})+P(2^\Theta|e_{1\overline{\wedge}i})=1$ as required in Equation (17) lead to

$$k_3\sum_{C\subseteq\Theta}\omega_i\widehat{m}_{C(1\overline{\wedge}i)}+k_3\omega_i\widehat{m}_{2^\Theta(1\overline{\wedge}i)}=1k_3=\frac{1}{\omega_i\left(\sum_{C\subseteq\Theta}\widehat{m}_{C(1\overline{\wedge}i)}+\widehat{m}_{2^\Theta(1\overline{\wedge}i)}\right)} \tag{A8-21}$$

Finally, putting Equation (A8-8), Equation (A8-18) and Equation (A8-21) into Equation (A8-7) leads to

$$\widetilde{m}_{2^\Theta(1\overline{\wedge}i)}=P(2^\Theta|e_{1\overline{\wedge}i})=k_3\omega_i\widehat{m}_{2^\Theta(1\overline{\wedge}i)}=\frac{\widehat{m}_{2^\Theta(1\overline{\wedge}i)}}{\sum_{C\subseteq\Theta}\widehat{m}_{C(1\overline{\wedge}i)}+\widehat{m}_{2^\Theta(1\overline{\wedge}i)}} \tag{A8-22}$$

$$\widetilde{m}_{\theta(1\overline{\wedge}i)}=P(\theta|e_{1\overline{\wedge}i})=k_3\omega_i\widehat{m}_{\theta(1\overline{\wedge}i)}=\frac{\widehat{m}_{\theta(1\overline{\wedge}i)}}{\sum_{C\subseteq\Theta}\widehat{m}_{C(1\overline{\wedge}i)}+\widehat{m}_{2^\Theta(1\overline{\wedge}i)}}\forall\theta\subseteq\Theta \tag{A8-23}$$

So, Equation (70) to Equation (72) also hold at l=i. Therefore, they hold for any l.

A9. Proof of Theorem 2 (Recursive MAKER rule)

Based on the likelihood principle, conditional probability $p(\theta|e_{1\overline{\wedge}L})$ measured in *SOPS* should be proportional to conditional probability $P(\theta|e_{1\overline{\wedge}L})$ measured in *ERPS* and generated from Lemma 2.1 at l=L-1, so that they hold the same evidential meanings. That is, the following equation should hold

$$p(\theta|e_{1\overline{\wedge}L}) = k_4 P(\theta|e_{1\overline{\wedge}L}) \qquad \forall\theta\subseteq\Theta \tag{A9-1}$$

where k_4 is a positive constant that does not change for any $\theta\subseteq\Theta$.

According to Definition 1, there should be $\sum_{D\subseteq\Theta} p(D|e_{1\overline{\wedge}L}) = 1$. From Equation (70) and Equation (A9-1), we then have

$$\sum_{D\subseteq\Theta} p(D|e_{1\overline{\wedge}L}) = \sum_{D\subseteq\Theta} k_4 P(D|e_{1\overline{\wedge}L}) = k_4 \sum_{D\subseteq\Theta} \frac{\widehat{m}_{D(1\overline{\wedge}L)}}{\sum_{C\subseteq\Theta}\widehat{m}_{C(1\overline{\wedge}L)} + \widehat{m}_{2^\Theta(1\overline{\wedge}L)}} = 1 \tag{A9-2}$$

So, we get

$$k_4 = \sum_{C\subseteq\Theta} \frac{\widehat{m}_{C(1\overline{\wedge}L)} + \widehat{m}_{2^\Theta(1\overline{\wedge}L)}}{\sum_{D\subseteq\Theta}\widehat{m}_{D(1\overline{\wedge}L)}} \tag{A9-3}$$

From Equation (70), putting Equation (A9-3) into Equation (A9-1) leads to

$$\begin{aligned} p(\theta|e_{1\overline{\wedge}L}) &= \frac{\sum_{C\subseteq\Theta}\widehat{m}_{C(1\overline{\wedge}L)} + \widehat{m}_{2^\Theta(1\overline{\wedge}L)}}{\sum_{D\subseteq\Theta}\widehat{m}_{D(1\overline{\wedge}L)}} P(\theta|e_{1\overline{\wedge}L}) \\ &= \sum_{C\subseteq\Theta} \frac{\widehat{m}_{C(1\overline{\wedge}L)} + \widehat{m}_{2^\Theta(1\overline{\wedge}L)}}{\sum_{D\subseteq\Theta}\widehat{m}_{D(1\overline{\wedge}L)}} \frac{\widehat{m}_{\theta(1\overline{\wedge}L)}}{\sum_{C\subseteq\Theta}\widehat{m}_{C(1\overline{\wedge}L)} + \widehat{m}_{2^\Theta(1\overline{\wedge}L)}} = \frac{\widehat{m}_{\theta(1\overline{\wedge}L)}}{\sum_{D\subseteq\Theta}\widehat{m}_{D(1\overline{\wedge}L)}} \qquad \forall\theta\subseteq\Theta \end{aligned} \tag{A9-4}$$

The probability of the residual support leftover after all pieces of evidence are combined is given by $\widetilde{m}_{2^\Theta(1\overline{\wedge}L)}$ that is calculated by Equation (A8-22) at $i = L$. The trustworthiness of the above combined results is then given by $1-\widetilde{m}_{2^\Theta(1\overline{\wedge}L)}$.

For $\theta = \emptyset$, there must be $p(\theta|e_{1\overline{\wedge}L}) = 0$ according to Definition 1.

A10. Proof of Corollary 2.1 (Additive MAKER algorithm)

Let $l_1 = l-1$ and $e_{1\overline{\wedge}l_1} = (e_1 \wedge \cdots \wedge e_{l-1})$ be the conjunction of the first l-1 pieces of evidence, and $e_{\theta(1\overline{\wedge}l_1)}$ be the evidential element that evidence $e_{1\overline{\wedge}l_1}$ points to state θ. If all pieces of evidence are exclusive of each other, from Equation (55) or Equation (56), by replacing e_{li} and e_{mj} with $e_{1\overline{\wedge}l_1}$ and e_l, respectively, we get $P\left((A|e_{1\overline{\wedge}l_1})\cap(B|e_l)\right) = p\left(\left(s_{A(1\overline{\wedge}l_1)}\middle|e_{e_{1\overline{\wedge}l_1}}\right)\cap(s_{Bl}|e_l)\right) = 0$ for any l and $A\cap B = \theta$. Equation (A8-18) and Equation (A8-19) then reduces to

$$P\left(\bigcup_{\substack{A\cap B=\theta\\ \theta\subseteq\Theta}}\left((A|e_{1\overline{\wedge}l_1})\cap(B|e_l)\right)\right) = \omega_l\left[\widetilde{m}_{\theta(1\overline{\wedge}l_1)}(1-r_l) + \widetilde{m}_{2^\Theta(1\overline{\wedge}l_1)}m_{\theta l}\right] = \omega_l\widehat{m}_{\theta(1\overline{\wedge}l)} \quad \forall\theta\subseteq\Theta \tag{A10-1}$$

$$\widehat{m}_{\theta(1\overline{\wedge}l)} = \left[\widetilde{m}_{\theta(1\overline{\wedge}l_1)}(1-r_l) + \widetilde{m}_{2^\Theta(1\overline{\wedge}l_1)}m_{\theta l}\right] \tag{A10-2}$$

The proof of Lemma 2.1 shows that Equation (71) reduces to Equation (A10-2). At L=2, from the proof of Lemma 2.1 and Theorem 2, we have for any $\theta\subseteq\Theta$

$$\begin{aligned} \widehat{m}_{\theta(1\overline{\wedge}2)} &= \widetilde{m}_{\theta 1}(1-r_2) + \widetilde{m}_{2^\Theta 1}m_{\theta 2} = m_{\theta 1}(1-r_2) + m_{2^\Theta 1}m_{\theta 2} \\ &= (1-r_2)m_{\theta 1} + (1-r_1)m_{\theta 2} = \sum_{i=1}^{2}\prod_{j=1, j\neq i}^{2}(1-r_j)m_{\theta i} \end{aligned} \tag{A10-3}$$

$$\widehat{m}_{2^\Theta(1\overline{\wedge}2)} = (1-r_2)\widetilde{m}_{2^\Theta 1} = (1-r_2)(1-r_1) = \prod_{j=1}^{2}(1-r_j) \tag{A10-4}$$

From Lemma 2.1 and Theorem 2, we then have

$$\widetilde{m}_{\theta(1\overline{\wedge}2)} = \frac{\widehat{m}_{\theta(1\overline{\wedge}2)}}{\sum_{C\subseteq\Theta}\widehat{m}_{C(1\overline{\wedge}2)} + \widehat{m}_{2^\Theta(1\overline{\wedge}2)}} = \frac{\sum_{i=1}^{2}\prod_{j=1,j\neq i}^{2}(1-r_j)m_{\theta i}}{\sum_{C\subseteq\Theta}\sum_{i=1}^{2}\prod_{j=1,j\neq i}^{2}(1-r_j)m_{Ci} + \prod_{j=1}^{2}(1-r_j)} \tag{A10-5}$$

$$\widetilde{m}_{2^\Theta(1\overline{\wedge}2)} = \frac{\widehat{m}_{2^\Theta(1\overline{\wedge}2)}}{\sum_{C\subseteq\Theta}\widehat{m}_{C(1\overline{\wedge}2)} + \widehat{m}_{2^\Theta(1\overline{\wedge}2)}} = \frac{\prod_{j=1}^{2}(1-r_j)}{\sum_{C\subseteq\Theta}\sum_{i=1}^{2}\prod_{j=1,j\neq i}^{2}(1-r_j)m_{Ci} + \prod_{j=1}^{2}(1-r_j)} \tag{A10-6}$$

$$p(\theta|e_{1\overline{\wedge}2}) = \frac{\widehat{m}_{\theta(1\overline{\wedge}2)}}{\sum_{C\subseteq\Theta}\widehat{m}_{C(1\overline{\wedge}2)}} = \frac{\sum_{i=1}^{2}\prod_{j=1,j\neq i}^{2}(1-r_j)m_{\theta i}}{\sum_{C\subseteq\Theta}\sum_{i=1}^{2}\prod_{j=1,j\neq i}^{2}(1-r_j)m_{Ci}} \tag{A10-7}$$

So, Equation (74) holds at L=2.
At L=3, from Equation (A10-2), Equation (A10-5) and Equation (A10-6), we have

$$\begin{aligned}
\widehat{m}_{\theta(1\overline{\wedge}3)} &= \widetilde{m}_{\theta(1\overline{\wedge}2)}(1-r_3) + \widetilde{m}_{2^\Theta(1\overline{\wedge}2)}m_{\theta 3} \\
&= \frac{\left(\sum_{i=1}^{2}\prod_{j=1,j\neq i}^{2}(1-r_j)m_{\theta i}\right)(1-r_3)}{\sum_{C\subseteq\Theta}\sum_{i=1}^{2}\prod_{j=1,j\neq i}^{2}(1-r_j)m_{Ci} + \prod_{j=1}^{2}(1-r_j)} + \frac{\left(\prod_{j=1}^{2}(1-r_j)\right)m_{\theta 3}}{\sum_{C\subseteq\Theta}\sum_{i=1}^{2}\prod_{j=1,j\neq i}^{2}(1-r_j)m_{Ci} + \prod_{j=1}^{2}(1-r_j)} \\
&= \frac{\left(\sum_{i=1}^{2}\prod_{j=1,j\neq i}^{2}(1-r_j)m_{\theta i}\right)(1-r_3) + \prod_{j=1}^{2}(1-r_j)m_{\theta 3}}{\sum_{C\subseteq\Theta}\sum_{i=1}^{2}\prod_{j=1,j\neq i}^{2}(1-r_j)m_{Ci} + \prod_{j=1}^{2}(1-r_j)} \\
&= \frac{\sum_{i=1}^{2}\prod_{j=1,j\neq i}^{3}(1-r_j)m_{\theta i} + \prod_{j=1,j\neq 3}^{3}(1-r_j)m_{\theta 3}}{\sum_{C\subseteq\Theta}\sum_{i=1}^{2}\prod_{j=1,j\neq i}^{2}(1-r_j)m_{Ci} + \prod_{j=1}^{2}(1-r_j)} \\
&= \frac{\sum_{i=1}^{3}\left(\prod_{j=1,j\neq i}^{3}(1-r_j)\right)m_{\theta i}}{\sum_{C\subseteq\Theta}\sum_{i=1}^{2}\prod_{j=1,j\neq i}^{2}(1-r_j)m_{Ci} + \prod_{j=1}^{2}(1-r_j)}
\end{aligned} \tag{A10-8}$$

$$\begin{aligned}
\widehat{m}_{2^\Theta(1\overline{\wedge}3)} &= (1-r_3)\widetilde{m}_{2^\Theta(1\overline{\wedge}2)} = \frac{(1-r_3)\prod_{j=1}^{2}(1-r_j)}{\sum_{C\subseteq\Theta}\sum_{i=1}^{2}\prod_{j=1,j\neq i}^{2}(1-r_j)m_{Ci} + \prod_{j=1}^{2}(1-r_j)} \\
&= \frac{\prod_{j=1}^{3}(1-r_j)}{\sum_{C\subseteq\Theta}\sum_{i=1}^{2}\prod_{j=1,j\neq i}^{2}(1-r_j)m_{Ci} + \prod_{j=1}^{2}(1-r_j)}
\end{aligned} \tag{A10-9}$$

$$\begin{aligned}
\widetilde{m}_{\theta(1\overline{\wedge}3)} &= \frac{\widehat{m}_{\theta(1\overline{\wedge}3)}}{\sum_{D\subseteq\Theta}\widehat{m}_{D(1\overline{\wedge}3)} + \widehat{m}_{2^\Theta(1\overline{\wedge}3)}} \\
&= \frac{\dfrac{\sum_{i=1}^{3}\left(\prod_{j=1,j\neq i}^{3}(1-r_j)\right)m_{\theta i}}{\sum_{C\subseteq\Theta}\sum_{i=1}^{2}\prod_{j=1,j\neq i}^{2}(1-r_j)m_{Ci} + \prod_{j=1}^{2}(1-r_j)}}{\sum_{D\subseteq\Theta}\dfrac{\sum_{i=1}^{3}\left(\prod_{j=1,j\neq i}^{3}(1-r_j)\right)m_{Di}}{\sum_{C\subseteq\Theta}\sum_{i=1}^{2}\prod_{j=1,j\neq i}^{2}(1-r_j)m_{Ci} + \prod_{j=1}^{2}(1-r_j)} + \dfrac{\prod_{j=1}^{3}(1-r_j)}{\sum_{C\subseteq\Theta}\sum_{i=1}^{2}\prod_{j=1,j\neq i}^{2}(1-r_j)m_{Ci} + \prod_{j=1}^{2}(1-r_j)}} \\
&= \frac{\sum_{i=1}^{3}\left(\prod_{j=1,j\neq i}^{3}(1-r_j)\right)m_{\theta i}}{\sum_{D\subseteq\Theta}\sum_{i=1}^{3}\left(\prod_{j=1,j\neq i}^{3}(1-r_j)\right)m_{Di} + \prod_{j=1}^{3}(1-r_j)}
\end{aligned} \tag{A10-10}$$

$$\begin{aligned}
\widetilde{m}_{2^\Theta(1\overline{\wedge}3)} &= \frac{\widehat{m}_{2^\Theta(1\overline{\wedge}3)}}{\sum_{D\subseteq\Theta}\widehat{m}_{D(1\overline{\wedge}3)} + \widehat{m}_{2^\Theta(1\overline{\wedge}3)}} \\
&= \frac{\dfrac{\prod_{j=1}^{3}(1-r_j)}{\sum_{C\subseteq\Theta}\sum_{i=1}^{2}\prod_{j=1,j\neq i}^{2}(1-r_j)m_{Ci} + \prod_{j=1}^{2}(1-r_j)}}{\sum_{D\subseteq\Theta}\dfrac{\sum_{i=1}^{3}\left(\prod_{j=1,j\neq i}^{3}(1-r_j)\right)m_{Di}}{\sum_{C\subseteq\Theta}\sum_{i=1}^{2}\prod_{j=1,j\neq i}^{2}(1-r_j)m_{Ci} + \prod_{j=1}^{2}(1-r_j)} + \dfrac{\prod_{j=1}^{3}(1-r_j)}{\sum_{C\subseteq\Theta}\sum_{i=1}^{2}\prod_{j=1,j\neq i}^{2}(1-r_j)m_{Ci} + \prod_{j=1}^{2}(1-r_j)}} \\
&= \frac{\prod_{j=1}^{3}(1-r_j)}{\sum_{D\subseteq\Theta}\sum_{i=1}^{3}\left(\prod_{j=1,j\neq i}^{3}(1-r_j)\right)m_{Di} + \prod_{j=1}^{3}(1-r_j)}
\end{aligned} \tag{A10-11}$$

$$p(\theta|e_{1\overline{\wedge}3}) = \frac{\widehat{m}_{\theta(1\overline{\wedge}3)}}{\sum_{D\subseteq\Theta}\widehat{m}_{D(1\overline{\wedge}3)}} = \frac{\dfrac{\sum_{i=1}^{3}\left(\prod_{j=1,j\neq i}^{3}(1-r_j)\right)m_{\theta i}}{\sum_{C\subseteq\Theta}\sum_{i=1}^{2}\left(\prod_{j=1,j\neq i}^{2}(1-r_j)\right)m_{Ci}+\prod_{j=1}^{2}(1-r_j)}}{\sum_{D\subseteq\Theta}\dfrac{\sum_{i=1}^{3}\left(\prod_{j=1,j\neq i}^{3}(1-r_j)\right)m_{Di}}{\sum_{C\subseteq\Theta}\sum_{i=1}^{2}\left(\prod_{j=1,j\neq i}^{2}(1-r_j)\right)m_{Ci}+\prod_{j=1}^{2}(1-r_j)}} = \frac{\sum_{i=1}^{3}\left(\prod_{j=1,j\neq i}^{3}(1-r_j)\right)m_{\theta i}}{\sum_{D\subseteq\Theta}\sum_{i=1}^{3}\left(\prod_{j=1,j\neq i}^{3}(1-r_j)\right)m_{Di}} \tag{A10-12}$$

So, Equation (74) holds at L=3.

From the formats of Equation (A10-8) and Equation (A10-9), we assume that the following Equations hold at $L = l_1 = l - 1$.

$$\widehat{m}_{\theta(1\overline{\wedge}l_1)} = \frac{\sum_{i=1}^{l_1}\left(\prod_{j=1,j\neq i}^{l_1}(1-r_j)\right)m_{\theta i}}{\sum_{C\subseteq\Theta}\sum_{i=1}^{l_1-1}\prod_{j=1,j\neq i}^{l_1-1}(1-r_j)m_{Ci}+\prod_{j=1}^{l_1-1}(1-r_j)} \tag{A10-13}$$

$$\widehat{m}_{2^\Theta(1\overline{\wedge}l_1)} = \frac{\prod_{j=1}^{l_1}(1-r_j)}{\sum_{C\subseteq\Theta}\sum_{i=1}^{l_1-1}\prod_{j=1,j\neq i}^{l_1-1}(1-r_j)m_{Ci}+\prod_{j=1}^{l_1-1}(1-r_j)} \tag{A10-14}$$

The above assumption leads to the following equations assumed to hold at $L = l_1$ as well:

$$\widetilde{m}_{\theta(1\overline{\wedge}l_1)} = \frac{\widehat{m}_{\theta(1\overline{\wedge}l_1)}}{\sum_{D\subseteq\Theta}\widehat{m}_{D(1\overline{\wedge}l_1)}+\widehat{m}_{2^\Theta(1\overline{\wedge}l_1)}} = \frac{\sum_{i=1}^{l_1}\left(\prod_{j=1,j\neq i}^{l_1}(1-r_j)\right)m_{\theta i}}{\sum_{D\subseteq\Theta}\sum_{i=1}^{l_1}\left(\prod_{j=1,j\neq i}^{l_1}(1-r_j)\right)m_{Di}+\prod_{j=1}^{l_1}(1-r_j)} \tag{A10-15}$$

$$\widetilde{m}_{2^\Theta(1\overline{\wedge}l_1)} = \frac{\widehat{m}_{2^\Theta(1\overline{\wedge}l_1)}}{\sum_{D\subseteq\Theta}\widehat{m}_{D(1\overline{\wedge}l_1)}+\widehat{m}_{2^\Theta(1\overline{\wedge}l_1)}} = \frac{\prod_{j=1}^{l_1}(1-r_j)}{\sum_{D\subseteq\Theta}\sum_{i=1}^{l_1}\left(\prod_{j=1,j\neq i}^{l_1}(1-r_j)\right)m_{Di}+\prod_{j=1}^{l_1}(1-r_j)}$$

$$p(\theta|e_{1\overline{\wedge}l_1}) = \frac{\widehat{m}_{\theta(1\overline{\wedge}l_1)}}{\sum_{C\subseteq\Theta}\widehat{m}_{C(1\overline{\wedge}l_1)}} = \frac{\sum_{i=1}^{l_1}\left(\prod_{j=1,j\neq i}^{l_1}(1-r_j)\right)m_{\theta i}}{\sum_{C\subseteq\Theta}\sum_{i=1}^{l_1}\left(\prod_{j=1,j\neq i}^{l_1}(1-r_j)\right)m_{Ci}} \tag{A10-16}$$

At L=l, from Lemma 2.1 and Theorem 2 with Equation (71) replaced by Equation (A10-2), we have

$$\begin{aligned}
\widehat{m}_{\theta(1\overline{\wedge}l)} &= \widetilde{m}_{\theta(1\overline{\wedge}l_1)}(1-r_l)+\widetilde{m}_{2^\Theta(1\overline{\wedge}l_1)}m_{\theta l}\\
&= \frac{\left(\sum_{i=1}^{l_1}\left(\prod_{j=1,j\neq i}^{l_1}(1-r_j)\right)m_{\theta i}\right)(1-r_l)}{\sum_{D\subseteq\Theta}\left(\sum_{i=1}^{l_1}\left(\prod_{j=1,j\neq i}^{l_1}(1-r_j)\right)m_{Di}\right)+\prod_{j=1}^{l_1}(1-r_j)} + \frac{\left(\prod_{j=1}^{l_1}(1-r_j)\right)m_{\theta l}}{\sum_{D\subseteq\Theta}\left(\sum_{i=1}^{l_1}\left(\prod_{j=1,j\neq i}^{l_1}(1-r_j)\right)m_{Di}\right)+\prod_{j=1}^{l_1}(1-r_j)}\\
&= \frac{\sum_{i=1}^{l_1}\left(\prod_{j=1,j\neq i}^{l}(1-r_j)\right)m_{\theta i}+\left(\prod_{j=1,j\neq l}^{l}(1-r_j)\right)m_{\theta l}}{\sum_{D\subseteq\Theta}\left(\sum_{i=1}^{l_1}\left(\prod_{j=1,j\neq i}^{l_1}(1-r_j)\right)m_{Di}\right)+\prod_{j=1}^{l_1}(1-r_j)}\\
&= \frac{\sum_{i=1}^{l}\left(\prod_{j=1,j\neq i}^{l}(1-r_j)\right)m_{\theta i}}{\sum_{D\subseteq\Theta}\left(\sum_{i=1}^{l_1}\left(\prod_{j=1,j\neq i}^{l_1}(1-r_j)\right)m_{Di}\right)+\prod_{j=1}^{l_1}(1-r_j)}
\end{aligned} \tag{A10-17}$$

$$\begin{aligned}
\widehat{m}_{2^\Theta(1\overline{\wedge}l)} &= (1-r_l)\widetilde{m}_{2^\Theta(1\overline{\wedge}l_1)}\\
&= \frac{(1-r_l)\prod_{j=1}^{l_1}(1-r_j)}{\sum_{D\subseteq\Theta}\sum_{i=1}^{l_1}\left(\prod_{j=1,j\neq i}^{l_1}(1-r_j)\right)m_{Di}+\prod_{j=1}^{l_1}(1-r_j)}\\
&= \frac{\prod_{j=1}^{l}(1-r_j)}{\sum_{D\subseteq\Theta}\sum_{i=1}^{l_1}\left(\prod_{j=1,j\neq i}^{l_1}(1-r_j)\right)m_{Di}+\prod_{j=1}^{l_1}(1-r_j)}
\end{aligned} \tag{A10-18}$$

From Equation (A10-17), Theorem 2 can then be re-written as follows:

$$p(\theta|e_{1\overline{\wedge}l}) = \frac{\widehat{m}_{\theta(1\overline{\wedge}l)}}{\sum_{C\subseteq\Theta}\widehat{m}_{C(1\overline{\wedge}l)}} = \frac{\dfrac{\sum_{i=1}^{l}\left(\prod_{j=1,j\neq i}^{l}(1-r_j)\right)m_{\theta i}}{\sum_{D\subseteq\Theta}\left(\sum_{i=1}^{l_1}\left(\prod_{j=1,j\neq i}^{l_1}(1-r_j)\right)m_{Di}\right)+\prod_{j=1}^{l_1}(1-r_j)}}{\sum_{C\subseteq\Theta}\dfrac{\sum_{i=1}^{l}\left(\prod_{j=1,j\neq i}^{l}(1-r_j)\right)m_{Ci}}{\sum_{D\subseteq\Theta}\left(\sum_{i=1}^{l_1}\left(\prod_{j=1,j\neq i}^{l_1}(1-r_j)\right)m_{Di}\right)+\prod_{j=1}^{l_1}(1-r_j)}}$$

$$= \frac{\sum_{i=1}^{l}\left(\prod_{j=1,j\neq i}^{l}(1-r_j)\right)m_{\theta i}}{\sum_{C\subseteq\Theta}\sum_{i=1}^{l}\left(\prod_{j=1,j\neq i}^{l}(1-r_j)\right)m_{Ci}} \tag{A10-19}$$

So, Equation (74) holds at $L=l$ as well. It therefore holds for any L.

Finally, we get the following equation for calculating the untrustworthiness of the combined results:

$$\widetilde{m}_{2^\Theta(1\overline{\wedge}l)} = \frac{\widehat{m}_{2^\Theta(1\overline{\wedge}l)}}{\sum_{C\subseteq\Theta}\widehat{m}_{C(1\overline{\wedge}l)}+\widehat{m}_{2^\Theta(1\overline{\wedge}l)}}$$

$$= \frac{\dfrac{\prod_{j=1}^{l}(1-r_j)}{\sum_{D\subseteq\Theta}\sum_{i=1}^{l_1}\left(\prod_{j=1,j\neq i}^{l_1}(1-r_j)\right)m_{Di}+\prod_{j=1}^{l_1}(1-r_j)}}{\sum_{C\in\Theta}\left(\dfrac{\sum_{i=1}^{l}\left(\prod_{j=1,j\neq i}^{l}(1-r_j)\right)m_{Ci}}{\sum_{D\subseteq\Theta}\left(\sum_{i=1}^{l_1}\left(\prod_{j=1,j\neq i}^{l_1}(1-r_j)\right)m_{Di}\right)+\prod_{j=1}^{l_1}(1-r_j)}\right)+\dfrac{\prod_{j=1}^{l}(1-r_j)}{\sum_{D\subseteq\Theta}\sum_{i=1}^{l_1}\left(\prod_{j=1,j\neq i}^{l_1}(1-r_j)\right)m_{Di}+\prod_{j=1}^{l_1}(1-r_j)}} \tag{A10-20}$$

$$= \frac{\prod_{j=1}^{l}(1-r_j)}{\sum_{C\in\Theta}\sum_{i=1}^{l}\left(\prod_{j=1,j\neq i}^{l}(1-r_j)\right)m_{Ci}+\prod_{j=1}^{l}(1-r_j)}$$

So, Equation (75) also holds at $L=l$. It therefore holds for any L.

A11. Proof of Corollary 2.2 (Recursive Independent MAKER algorithm)

Let $l_1 = l-1$ and $e_{1\overline{\wedge}l_1} = (e_1 \wedge \cdots \wedge e_{l-1})$ be the conjunction of the first l-1 pieces of evidence, and $e_{\theta(1\overline{\wedge}l_1)}$ be the evidential element that evidence $e_{1\overline{\wedge}l_1}$ points to state θ. Since all pieces of evidence are mutually independent, from Equation (20), Equation (22), Equation (23) and Equation (57), the joint probability in Equation (A8-10) of Appendix A8 becomes

$$P((A|e_{1\overline{\wedge}l_1})\cap(B|e_l)) = p((s_{A(1\overline{\wedge}l_1)}|e_{1\overline{\wedge}l_1})\cap(s_{Bl}|e_l)) = p(s_{A(1\overline{\wedge}l_1)}|e_{1\overline{\wedge}l_1})p(s_{Bl}|e_l) = \widetilde{m}_{A(1\overline{\wedge}l_1)}\omega_l m_{Bl} \tag{A11-1}$$

Putting Equation (A11-1) into Equation (A8-10) leads to

$$P\left(\bigcup_{\substack{A\cap B=\theta\\ \theta\subseteq\Theta}}((A|e_{1\overline{\wedge}l_1})\cap(B|e_l))\right) = \omega_l[\widetilde{m}_{\theta(1\overline{\wedge}l_1)}(1-r_l)+\widetilde{m}_{2^\Theta(1\overline{\wedge}l_1)}m_{\theta l}] + \omega_l\sum_{\substack{A\cap B=\theta\\ A,\ B\subseteq\Theta}}\widetilde{m}_{A(1\overline{\wedge}l_1)}m_{Bl} = \omega_l\widehat{m}_{\theta(1\overline{\wedge}l)} \tag{A11-2}$$

$$\widehat{m}_{\theta(1\overline{\wedge}l)} = [\widetilde{m}_{\theta(1\overline{\wedge}l_1)}(1-r_l)+\widetilde{m}_{2^\Theta(1\overline{\wedge}l_1)}m_{\theta l}] + \sum_{\substack{A\cap B=\theta\\ A,\ B\subseteq\Theta}}\widetilde{m}_{A(1\overline{\wedge}l_1)}m_{Bl} \tag{A11-3}$$

Replacing Equation (A8-18) and Equation (A8-19) by Equation (A11-2) and Equation (A11-3), we can prove Corollary 2.2 in the same way as for proving Lemma 2.1.

A12. Proof of Corollary 2.3 (Multiplicative MAKER algorithm)

Let $l_1 = l-1$ and $e_{1\overline{\wedge}l_1} = (e_1 \wedge \cdots \wedge e_{l-1})$ be the conjunction of the first l-1 pieces of evidence, and $e_{\theta(1\overline{\wedge}l_1)}$ be the evidential element that evidence $e_{1\overline{\wedge}l_1}$ points to state θ. If all L pieces of evidence are independent of each other and basic probabilities are assigned to singleton system states and the system space only, Equation (A11-3) reduces to

$$\begin{aligned}
\widehat{m}_{\theta(1\overline{\wedge}l)} &= [\widetilde{m}_{\theta(1\overline{\wedge}l_1)}(1-r_l) + \widetilde{m}_{2^\Theta(1\overline{\wedge}l_1)}m_{\theta l}] + \widetilde{m}_{\theta(1\overline{\wedge}l_1)}m_{\theta l} + \widetilde{m}_{\theta(1\overline{\wedge}l_1)}m_{\Theta l} + \widetilde{m}_{\Theta(1\overline{\wedge}l_1)}m_{\theta l} \\
&= \widetilde{m}_{\theta(1\overline{\wedge}l_1)}m_{\theta l} + \widetilde{m}_{\theta(1\overline{\wedge}l_1)}(m_{\Theta l} + (1-r_l)) + (\widetilde{m}_{\Theta(1\overline{\wedge}l_1)} + \widetilde{m}_{2^\Theta(1\overline{\wedge}l_1)})m_{\theta l} \\
&= \widetilde{m}_{\theta(1\overline{\wedge}l_1)}(m_{\theta l} + m_{\Theta l} + (1-r_l)) + (\widetilde{m}_{\Theta(1\overline{\wedge}l_1)} + \widetilde{m}_{2^\Theta(1\overline{\wedge}l_1)})m_{\theta l} \\
&= \widetilde{m}_{\theta(1\overline{\wedge}l_1)}(m_{\theta l} + m_{\Theta l} + (1-r_l)) + (\widetilde{m}_{\Theta(1\overline{\wedge}l_1)} + \widetilde{m}_{2^\Theta(1\overline{\wedge}l_1)})(m_{\theta l} + m_{\Theta l} + (1-r_l)) \\
&\quad -(\widetilde{m}_{\Theta(1\overline{\wedge}l_1)} + \widetilde{m}_{2^\Theta(1\overline{\wedge}l_1)})(m_{\Theta l} + (1-r_l)) \\
&= (\widetilde{m}_{\theta(1\overline{\wedge}l_1)} + \widetilde{m}_{\Theta(1\overline{\wedge}l_1)} + \widetilde{m}_{2^\Theta(1\overline{\wedge}l_1)})(m_{\theta l} + m_{\Theta l} + (1-r_l)) - (\widetilde{m}_{\Theta(1\overline{\wedge}l_1)} + \widetilde{m}_{2^\Theta(1\overline{\wedge}l_1)})(m_{\Theta l} + (1-r_l))
\end{aligned} \tag{A12-1}$$

$$\begin{aligned}
\widehat{m}_{\Theta(1\overline{\wedge}l)} &= [\widetilde{m}_{\Theta(1\overline{\wedge}l_1)}(1-r_l) + \widetilde{m}_{2^\Theta(1\overline{\wedge}l_1)}m_{\Theta l}] + \widetilde{m}_{\Theta(1\overline{\wedge}l_1)}m_{\Theta l} \\
&= (\widetilde{m}_{\Theta(1\overline{\wedge}l_1)} + \widetilde{m}_{2^\Theta(1\overline{\wedge}l_1)})m_{\Theta l} + \widetilde{m}_{\Theta(1\overline{\wedge}l_1)}(1-r_l) \\
&= (\widetilde{m}_{\Theta(1\overline{\wedge}l_1)} + \widetilde{m}_{2^\Theta(1\overline{\wedge}l_1)})m_{\Theta l} + (\widetilde{m}_{\Theta(1\overline{\wedge}l_1)} + \widetilde{m}_{2^\Theta(1\overline{\wedge}l_1)})(1-r_l) \\
&\quad -\widetilde{m}_{2^\Theta(1\overline{\wedge}l_1)}(1-r_l) \\
&= (\widetilde{m}_{\Theta(1\overline{\wedge}l_1)} + \widetilde{m}_{2^\Theta(1\overline{\wedge}l_1)})(m_{\Theta l} + (1-r_l)) - \widetilde{m}_{2^\Theta(1\overline{\wedge}l_1)}(1-r_l)
\end{aligned} \tag{A12-2}$$

$$\widehat{m}_{2^\Theta(1\overline{\wedge}l)} = \widetilde{m}_{2^\Theta(1\overline{\wedge}l_1)}(1-r_l) \tag{A12-3}$$

At $L = 2$, from Equation (A12-1) to Equation (A12-3) we get

$$\begin{aligned}
\widehat{m}_{\theta(1\overline{\wedge}2)} &= (\widetilde{m}_{\theta 1} + \widetilde{m}_{\Theta 1} + \widetilde{m}_{2^\Theta 1})(m_{\theta 2} + m_{\Theta 2} + (1-r_2)) \\
&\quad -(\widetilde{m}_{\Theta 1} + \widetilde{m}_{2^\Theta 1})(m_{\Theta 2} + (1-r_2)) \\
&= (m_{\theta 1} + m_{\Theta 1} + (1-r_1))(m_{\theta 2} + m_{\Theta 2} + (1-r_2)) \\
&\quad -(m_{\Theta 1} + (1-r_1))(m_{\Theta 2} + (1-r_2)) \\
&= \prod_{i=1}^{2}(m_{\theta i} + m_{\Theta i} + (1-r_i)) - \prod_{i=1}^{2}(m_{\Theta i} + (1-r_i))
\end{aligned} \tag{A12-4}$$

$$\begin{aligned}
\widehat{m}_{\Theta(1\overline{\wedge}2)} &= (\widetilde{m}_{\Theta 1} + \widetilde{m}_{2^\Theta 1})(m_{\Theta 2} + (1-r_2)) - \widetilde{m}_{2^\Theta 1}(1-r_2) \\
&= (m_{\Theta 1} + m_{2^\Theta 1})(m_{\Theta 2} + (1-r_2)) - m_{2^\Theta 1}(1-r_2) \\
&= (m_{\Theta 1} + (1-r_1))(m_{\Theta 2} + (1-r_2)) - (1-r_1)(1-r_2) \\
&= \prod_{i=1}^{2}(m_{\Theta i} + (1-r_i)) - \prod_{i=1}^{2}(1-r_i)
\end{aligned} \tag{A12-5}$$

$$\widehat{m}_{2^\Theta(1\overline{\wedge}2)} = (1-r_2)\widetilde{m}_{2^\Theta 1} = (1-r_2)m_{2^\Theta 1} = (1-r_2)(1-r_1) = \prod_{i=1}^{2}(1-r_i) \tag{A12-6}$$

From Lemma 2.1, Theorem 2 and Equation (A12-4) to Equation (A12-6), we therefore get

$$\begin{aligned}
k_{1\overline{\wedge}2} &= \sum_{C\subseteq\Theta}\widehat{m}_{C(1\overline{\wedge}2)} + \widehat{m}_{2^\Theta(1\overline{\wedge}2)} = \sum_{C\in\Theta}\widehat{m}_{C(1\overline{\wedge}2)} + \widehat{m}_{\Theta(1\overline{\wedge}2)} + \widehat{m}_{2^\Theta(1\overline{\wedge}2)} \\
&= \sum_{C\in\Theta}\left(\prod_{i=1}^{2}(m_{Ci} + m_{\Theta i} + (1-r_i)) - \prod_{i=1}^{2}(m_{\Theta i} + (1-r_i))\right) \\
&\quad + \left(\prod_{i=1}^{2}(m_{\Theta i} + (1-r_i)) - \prod_{i=1}^{2}(1-r_i)\right) + \prod_{i=1}^{2}(1-r_i) \\
&= \sum_{C\in\Theta}\prod_{i=1}^{2}(m_{Ci} + m_{\Theta i} + (1-r_i)) - N\prod_{i=1}^{2}(m_{\Theta i} + (1-r_i)) \\
&\quad + \prod_{i=1}^{2}(m_{\Theta i} + (1-r_i)) \\
&= \sum_{C\in\Theta}\prod_{i=1}^{2}(m_{Ci} + m_{\Theta i} + (1-r_i)) - (N-1)\prod_{i=1}^{2}(m_{\Theta i} + (1-r_i))
\end{aligned} \tag{A12-7}$$

$$\widetilde{m}_{\theta(1\overline{\wedge}2)} = \frac{\widehat{m}_{\theta(1\overline{\wedge}2)}}{\sum_{C\subseteq\Theta}\widehat{m}_{C(1\overline{\wedge}2)} + \widehat{m}_{2^\Theta(1\overline{\wedge}2)}} = \frac{\prod_{i=1}^{2}(m_{\theta i} + m_{\Theta i} + (1-r_i)) - \prod_{i=1}^{2}(m_{\Theta i} + (1-r_i))}{k_{1\overline{\wedge}2}} \tag{A12-8}$$

$$\widetilde{m}_{\Theta(1\overline{\wedge}2)} = \frac{\widehat{m}_{\Theta(1\overline{\wedge}2)}}{\sum_{C\subseteq\Theta}\widehat{m}_{C(1\overline{\wedge}2)} + \widehat{m}_{2^\Theta(1\overline{\wedge}2)}} = \frac{\prod_{i=1}^{2}(m_{\Theta i} + (1-r_i)) - \prod_{i=1}^{2}(1-r_i)}{k_{1\overline{\wedge}2}} \tag{A12-9}$$

$$\widetilde{m}_{2^\Theta(1\overline{\wedge}2)} = \frac{\widehat{m}_{2^\Theta(1\overline{\wedge}2)}}{\sum_{C\subseteq\Theta}\widehat{m}_{C(1\overline{\wedge}2)} + \widehat{m}_{2^\Theta(1\overline{\wedge}2)}} = \frac{\prod_{i=1}^{2}(1-r_i)}{k_{1\overline{\wedge}2}} \tag{A12-10}$$

$$p(\theta|e_{1\overline{\wedge}2}) = \frac{\widehat{m}_{\theta(1\overline{\wedge}2)}}{\sum_{C\subseteq\Theta}\widehat{m}_{C(1\overline{\wedge}2)}} = \frac{\widehat{m}_{\theta(1\overline{\wedge}2)}}{\sum_{C\in\Theta}\widehat{m}_{C(1\overline{\wedge}2)} + \widehat{m}_{\Theta(1\overline{\wedge}2)}}$$
$$= \frac{\prod_{i=1}^{2}(m_{\theta i} + m_{\Theta i} + (1-r_i)) - \prod_{i=1}^{2}(m_{\Theta i} + (1-r_i))}{\sum_{C\in\Theta}\left(\prod_{i=1}^{2}(m_{Ci} + m_{\Theta i} + (1-r_i)) - \prod_{i=1}^{2}(m_{\Theta i} + (1-r_i))\right) + \left(\prod_{i=1}^{2}(m_{\Theta i} + (1-r_i)) - \prod_{i=1}^{2}(1-r_i)\right)}$$
$$= \frac{\prod_{i=1}^{2}(m_{\theta i} + m_{\Theta i} + (1-r_i)) - \prod_{i=1}^{2}(m_{\Theta i} + (1-r_i))}{\sum_{C\in\Theta}\prod_{i=1}^{2}(m_{Ci} + m_{\Theta i} + (1-r_i)) - N\prod_{i=1}^{2}(m_{\Theta i} + (1-r_i)) + \left(\prod_{i=1}^{2}(m_{\Theta i} + (1-r_i)) - \prod_{i=1}^{2}(1-r_i)\right)}$$
$$= \frac{\prod_{i=1}^{2}(m_{\theta i} + m_{\Theta i} + (1-r_i)) - \prod_{i=1}^{2}(m_{\Theta i} + (1-r_i))}{\sum_{C\in\Theta}\prod_{i=1}^{2}(m_{Ci} + m_{\Theta i} + (1-r_i)) - (N-1)\prod_{i=1}^{2}(m_{\Theta i} + (1-r_i)) - \prod_{i=1}^{2}(1-r_i)} \tag{A12-11}$$

$$p(\Theta|e_{1\overline{\wedge}2}) = \frac{\widehat{m}_{\Theta(1\overline{\wedge}2)}}{\sum_{C\subseteq\Theta}\widehat{m}_{C(1\overline{\wedge}2)}} = \frac{\widehat{m}_{\Theta(1\overline{\wedge}2)}}{\sum_{C\in\Theta}\widehat{m}_{C(1\overline{\wedge}2)} + \widehat{m}_{\Theta(1\overline{\wedge}2)}}$$
$$= \frac{\prod_{i=1}^{2}(m_{\Theta i} + (1-r_i)) - \prod_{i=1}^{2}(1-r_i)}{\sum_{C\in\Theta}\left(\prod_{i=1}^{2}(m_{Ci} + m_{\Theta i} + (1-r_i)) - \prod_{i=1}^{2}(m_{\Theta i} + (1-r_i))\right) + \left(\prod_{i=1}^{2}(m_{\Theta i} + (1-r_i)) - \prod_{i=1}^{2}(1-r_i)\right)}$$
$$= \frac{\prod_{i=1}^{2}(m_{\Theta i} + (1-r_i)) - \prod_{i=1}^{2}(1-r_i)}{\sum_{C\in\Theta}\prod_{i=1}^{2}(m_{Ci} + m_{\Theta i} + (1-r_i)) - (N-1)\prod_{i=1}^{2}(m_{\Theta i} + (1-r_i)) - \prod_{i=1}^{2}(1-r_i)} \tag{A12-12}$$

So, Equation (77) holds at L=2.

At $L = 3$, from Equation (A12-7) to Equation (A12-10) we have

$$\left(\widetilde{m}_{\theta(1\overline{\wedge}2)} + \widetilde{m}_{\Theta(1\overline{\wedge}2)} + \widetilde{m}_{2^\Theta(1\overline{\wedge}2)}\right)$$
$$= \frac{\prod_{i=1}^{2}(m_{\theta i} + m_{\Theta i} + (1-r_i)) - \prod_{i=1}^{2}(m_{\Theta i} + (1-r_i))}{k_{1\overline{\wedge}2}} + \frac{\prod_{i=1}^{2}(m_{\Theta i} + (1-r_i)) - \prod_{i=1}^{2}(1-r_i)}{k_{1\overline{\wedge}2}}$$
$$+ \frac{\prod_{i=1}^{2}(1-r_i)}{k_{1\overline{\wedge}2}}$$
$$= \frac{\prod_{i=1}^{2}(m_{\theta i} + m_{\Theta i} + (1-r_i))}{k_{1\overline{\wedge}2}} \tag{A12-13}$$

$$\left(\widetilde{m}_{\Theta(1\overline{\wedge}2)} + \widetilde{m}_{2^\Theta(1\overline{\wedge}2)}\right) = \frac{\prod_{i=1}^{2}(m_{\Theta i} + (1-r_i)) - \prod_{i=1}^{2}(1-r_i)}{k_{1\overline{\wedge}2}} + \frac{\prod_{i=1}^{2}(1-r_i)}{k_{1\overline{\wedge}2}} = \frac{\prod_{i=1}^{2}(m_{\Theta i} + (1-r_i))}{k_{1\overline{\wedge}2}} \tag{A12-14}$$

From Equation (A12-1) to Equation (A12-3) and the above two equations, we then get

$$\widehat{m}_{\theta(1\overline{\wedge}3)} = \left(\widetilde{m}_{\theta(1\overline{\wedge}2)} + \widetilde{m}_{\Theta(1\overline{\wedge}2)} + \widetilde{m}_{2^\Theta(1\overline{\wedge}2)}\right)(m_{\theta 3} + m_{\Theta 3} + (1-r_3))$$
$$-\left(\widetilde{m}_{\Theta(1\overline{\wedge}2)} + \widetilde{m}_{2^\Theta(1\overline{\wedge}2)}\right)(m_{\Theta 3} + (1-r_3))$$
$$= \frac{\prod_{i=1}^{2}(m_{\theta i} + m_{\Theta i} + (1-r_i))}{k_{1\overline{\wedge}2}}(m_{\theta 3} + m_{\Theta 3} + (1-r_3)) - \frac{\prod_{i=1}^{2}(m_{\Theta i} + (1-r_i))}{k_{1\overline{\wedge}2}}(m_{\Theta 3} + (1-r_3))$$
$$= \frac{\prod_{i=1}^{3}(m_{\theta i} + m_{\Theta i} + (1-r_i))}{k_{1\overline{\wedge}2}} - \frac{\prod_{i=1}^{3}(m_{\Theta i} + (1-r_i))}{k_{1\overline{\wedge}2}} \tag{A12-15}$$

$$\widehat{m}_{\Theta(1\overline{\wedge}3)} = \left(\widetilde{m}_{\Theta(1\overline{\wedge}2)} + \widetilde{m}_{2^\Theta(1\overline{\wedge}2)}\right)(m_{\Theta 3} + (1-r_3)) - \widetilde{m}_{2^\Theta(1\overline{\wedge}2)}(1-r_3)$$
$$= \frac{\prod_{i=1}^{2}(m_{\Theta i} + (1-r_i))}{k_{1\overline{\wedge}2}}(m_{\Theta 3} + (1-r_3)) - \frac{\prod_{i=1}^{2}(1-r_i)}{k_{1\overline{\wedge}2}}(1-r_3)$$
$$= \frac{\prod_{i=1}^{3}(m_{\Theta i} + (1-r_i))}{k_{1\overline{\wedge}2}} - \frac{\prod_{i=1}^{3}(1-r_i)}{k_{1\overline{\wedge}2}} \tag{A12-16}$$

$$\widehat{m}_{2^\Theta(1\overline{\wedge}3)} = \widetilde{m}_{2^\Theta(1\overline{\wedge}2)}(1-r_3) = \frac{\prod_{i=1}^{2}(1-r_i)}{k_{1\overline{\wedge}2}}(1-r_3) = \frac{\prod_{i=1}^{3}(1-r_i)}{k_{1\overline{\wedge}2}} \tag{A12-17}$$

Following the formats of Equation (A12-7) and Equation (A12-15) to Equation (A12-17), we assume that the following equations hold at L=l_1.

$$k_{1\overline{\wedge}(l_1-1)} = \sum_{C\in\Theta}\prod_{i=1}^{l_1-1}(m_{Ci} + m_{\Theta i} + (1-r_i)) - (N-1)\prod_{i=1}^{l_1-1}(m_{\Theta i} + (1-r_i)) \tag{A12-18}$$

$$\hat{m}_{\theta(1\overline{\wedge}l_1)} = \frac{\prod_{i=1}^{l_1}(m_{\theta i} + m_{\Theta i} + (1-r_i))}{k_{1\overline{\wedge}(l_1-1)}} - \frac{\prod_{i=1}^{l_1}(m_{\Theta i} + (1-r_i))}{k_{1\overline{\wedge}(l_1-1)}} \tag{A12-19}$$

$$\hat{m}_{\Theta(1\overline{\wedge}l_1)} = \frac{\prod_{i=1}^{l_1}(m_{\Theta i} + (1-r_i))}{k_{1\overline{\wedge}(l_1-1)}} - \frac{\prod_{i=1}^{l_1}(1-r_i)}{k_{1\overline{\wedge}(l_1-1)}} \tag{A12-20}$$

$$\hat{m}_{2^\Theta(1\overline{\wedge}l_1)} = \frac{\prod_{i=1}^{l_1}(1-r_i)}{k_{1\overline{\wedge}(l_1-1)}} \tag{A12-21}$$

From Lemma 2.1, Theorem 2 and Equation (A12-18) to Equation (A12-21), we then get

$$\begin{aligned}
\tilde{m}_{\theta(1\overline{\wedge}l_1)} &= \frac{\hat{m}_{\theta(1\overline{\wedge}l_1)}}{\sum_{C\subseteq\Theta}\hat{m}_{C(1\overline{\wedge}l_1)} + \hat{m}_{2^\Theta(1\overline{\wedge}l_1)}} = \frac{\hat{m}_{\theta(1\overline{\wedge}l_1)}}{\sum_{C\in\Theta}\hat{m}_{C(1\overline{\wedge}l_1)} + \hat{m}_{\Theta(1\overline{\wedge}l_1)} + \hat{m}_{2^\Theta(1\overline{\wedge}l_1)}} \\
&= \frac{\frac{\prod_{i=1}^{l_1}(m_{\theta i} + m_{\Theta i} + (1-r_i))}{k_{1\overline{\wedge}(l_1-1)}} - \frac{\prod_{i=1}^{l_1}(m_{\Theta i} + (1-r_i))}{k_{1\overline{\wedge}(l_1-1)}}}{\sum_{C\in\Theta}\left(\frac{\prod_{i=1}^{l_1}(m_{Ci} + m_{\Theta i} + (1-r_i))}{k_{1\overline{\wedge}(l_1-1)}} - \frac{\prod_{i=1}^{l_1}(m_{\Theta i} + (1-r_i))}{k_{1\overline{\wedge}(l_1-1)}}\right) + \left(\frac{\prod_{i=1}^{l_1}(m_{\Theta i} + (1-r_i))}{k_{1\overline{\wedge}(l_1-1)}} - \frac{\prod_{i=1}^{l_1}(1-r_i)}{k_{1\overline{\wedge}(l_1-1)}}\right) + \frac{\prod_{i=1}^{l_1}(1-r_i)}{k_{1\overline{\wedge}(l_1-1)}}} \\
&= \frac{\prod_{i=1}^{l_1}(m_{\theta i} + m_{\Theta i} + (1-r_i)) - \prod_{i=1}^{l_1}(m_{\Theta i} + (1-r_i))}{\sum_{C\in\Theta}\left(\prod_{i=1}^{l_1}(m_{Ci} + m_{\Theta i} + (1-r_i)) - \prod_{i=1}^{l_1}(m_{\Theta i} + (1-r_i))\right) + \left(\prod_{i=1}^{l_1}(m_{\Theta i} + (1-r_i)) - \prod_{i=1}^{l_1}(1-r_i)\right) + \prod_{i=1}^{l_1}(1-r_i)} \\
&= \frac{\prod_{i=1}^{l_1}(m_{\theta i} + m_{\Theta i} + (1-r_i)) - \prod_{i=1}^{l_1}(m_{\Theta i} + (1-r_i))}{\sum_{C\in\Theta}\prod_{i=1}^{l_1}(m_{Ci} + m_{\Theta i} + (1-r_i)) - N\prod_{i=1}^{l_1}(m_{\Theta i} + (1-r_i)) + \prod_{i=1}^{l_1}(m_{\Theta i} + (1-r_i))} \\
&= \frac{\prod_{i=1}^{l_1}(m_{\theta i} + m_{\Theta i} + (1-r_i)) - \prod_{i=1}^{l_1}(m_{\Theta i} + (1-r_i))}{\sum_{C\in\Theta}\prod_{i=1}^{l_1}(m_{Ci} + m_{\Theta i} + (1-r_i)) - (N-1)\prod_{i=1}^{l_1}(m_{\Theta i} + (1-r_i))} \\
&= \frac{\prod_{i=1}^{l_1}(m_{\theta i} + m_{\Theta i} + (1-r_i)) - \prod_{i=1}^{l_1}(m_{\Theta i} + (1-r_i))}{k_{1\overline{\wedge}l_1}}
\end{aligned} \tag{A12-22}$$

$$k_{1\overline{\wedge}l_1} = \sum_{C\in\Theta}\prod_{i=1}^{l_1}(m_{Ci} + m_{\Theta i} + (1-r_i)) - (N-1)\prod_{i=1}^{l_1}(m_{\Theta i} + (1-r_i)) \tag{A12-23}$$

Similarly, we get

$$\tilde{m}_{\Theta(1\overline{\wedge}l_1)} = \frac{\hat{m}_{\Theta(1\overline{\wedge}l_1)}}{\sum_{C\subseteq\Theta}\hat{m}_{C(1\overline{\wedge}l_1)} + \hat{m}_{2^\Theta(1\overline{\wedge}l_1)}} = \frac{\prod_{i=1}^{l_1}(m_{\Theta i} + (1-r_i)) - \prod_{i=1}^{l_1}(1-r_i)}{k_{1\overline{\wedge}l_1}} \tag{A12-24}$$

$$\tilde{m}_{2^\Theta(1\overline{\wedge}l_1)} = \frac{\hat{m}_{2^\Theta(1\overline{\wedge}l_1)}}{\sum_{C\subseteq\Theta}\hat{m}_{C(1\overline{\wedge}l_1)} + \hat{m}_{2^\Theta(1\overline{\wedge}l_1)}} = \frac{\prod_{i=1}^{l_1}(1-r_i)}{k_{1\overline{\wedge}l_1}} \tag{A12-25}$$

From Equation (A12-18) to Equation (A12-21), we also get

$$\begin{aligned}
p(\theta|e_{1\overline{\wedge}l_1}) &= \frac{\hat{m}_{\theta(1\overline{\wedge}l_1)}}{\sum_{C\subseteq\Theta}\hat{m}_{C(1\overline{\wedge}l_1)}} = \frac{\hat{m}_{\theta(1\overline{\wedge}l_1)}}{\sum_{C\in\Theta}\hat{m}_{C(1\overline{\wedge}l_1)} + \hat{m}_{\Theta(1\overline{\wedge}l_1)}} \\
&= \frac{\frac{\prod_{i=1}^{l_1}(m_{\theta i} + m_{\Theta i} + (1-r_i))}{k_{1\overline{\wedge}(l_1-1)}} - \frac{\prod_{i=1}^{l_1}(m_{\Theta i} + (1-r_i))}{k_{1\overline{\wedge}(l_1-1)}}}{\sum_{C\in\Theta}\left(\frac{\prod_{i=1}^{l_1}(m_{Ci} + m_{\Theta i} + (1-r_i))}{k_{1\overline{\wedge}(l_1-1)}} - \frac{\prod_{i=1}^{l_1}(m_{\Theta i} + (1-r_i))}{k_{1\overline{\wedge}(l_1-1)}}\right) + \left(\frac{\prod_{i=1}^{l_1}(m_{\Theta i} + (1-r_i))}{k_{1\overline{\wedge}(l_1-1)}} - \frac{\prod_{i=1}^{l_1}(1-r_i)}{k_{1\overline{\wedge}(l_1-1)}}\right)} \\
&= \frac{\prod_{i=1}^{l_1}(m_{\theta i} + m_{\Theta i} + (1-r_i)) - \prod_{i=1}^{l_1}(m_{\Theta i} + (1-r_i))}{\sum_{C\in\Theta}\left(\prod_{i=1}^{l_1}(m_{Ci} + m_{\Theta i} + (1-r_i)) - \prod_{i=1}^{l_1}(m_{\Theta i} + (1-r_i))\right) + \left(\prod_{i=1}^{l_1}(m_{\Theta i} + (1-r_i)) - \prod_{i=1}^{l_1}(1-r_i)\right)} \\
&= \frac{\prod_{i=1}^{l_1}(m_{\theta i} + m_{\Theta i} + (1-r_i)) - \prod_{i=1}^{l_1}(m_{\Theta i} + (1-r_i))}{\sum_{C\in\Theta}\prod_{i=1}^{l_1}(m_{Ci} + m_{\Theta i} + (1-r_i)) - N\prod_{i=1}^{l_1}(m_{\Theta i} + (1-r_i)) + \prod_{i=1}^{l_1}(m_{\Theta i} + (1-r_i)) - \prod_{i=1}^{l_1}(1-r_i)} \\
&= \frac{\prod_{i=1}^{l_1}(m_{\theta i} + m_{\Theta i} + (1-r_i)) - \prod_{i=1}^{l_1}(m_{\Theta i} + (1-r_i))}{\sum_{C\in\Theta}\prod_{i=1}^{l_1}(m_{Ci} + m_{\Theta i} + (1-r_i)) - (N-1)\prod_{i=1}^{l_1}(m_{\Theta i} + (1-r_i)) - \prod_{i=1}^{l_1}(1-r_i)}
\end{aligned} \tag{A12-26}$$

$$p(\Theta|e_{1\bar{\wedge}l_1}) = \frac{\widehat{m}_{\Theta(1\bar{\wedge}l_1)}}{\sum_{C\subseteq\Theta}\widehat{m}_{C(1\bar{\wedge}l_1)}} = \frac{\widehat{m}_{\Theta(1\bar{\wedge}l_1)}}{\sum_{C\in\Theta}\widehat{m}_{C(1\bar{\wedge}l_1)} + \widehat{m}_{\Theta(1\bar{\wedge}l_1)}}$$
$$= \frac{\frac{\prod_{i=1}^{l_1}(m_{\Theta i}+(1-r_i))}{k_{1\bar{\wedge}(l_1-1)}} - \frac{\prod_{i=1}^{l_1}(1-r_i)}{k_{1\bar{\wedge}(l_1-1)}}}{\sum_{C\in\Theta}\left(\frac{\prod_{i=1}^{l_1}(m_{Ci}+m_{\Theta i}+(1-r_i))}{k_{1\bar{\wedge}(l_1-1)}} - \frac{\prod_{i=1}^{l_1}(m_{\Theta i}+(1-r_i))}{k_{1\bar{\wedge}(l_1-1)}}\right) + \left(\frac{\prod_{i=1}^{l_1}(m_{\Theta i}+(1-r_i))}{k_{1\bar{\wedge}(l_1-1)}} - \frac{\prod_{i=1}^{l_1}(1-r_i)}{k_{1\bar{\wedge}(l_1-1)}}\right)}$$
$$= \frac{\prod_{i=1}^{l_1}(m_{\Theta i}+(1-r_i)) - \prod_{i=1}^{l_1}(1-r_i)}{\sum_{C\in\Theta}\prod_{i=1}^{l_1}(m_{Ci}+m_{\Theta i}+(1-r_i)) - (N-1)\prod_{i=1}^{l_1}(m_{\Theta i}+(1-r_i)) - \prod_{i=1}^{l_1}(1-r_i)} \quad \text{(A12-27)}$$

The above results show that the assumption of Equation (A12-18) to Equation (A12-21) leads to Equation (77) holding at $L=l_1$ as well.

At $L=l$, from Equation (A12-22) to Equation (A12 -25), we have

$$(\widetilde{m}_{\theta(1\bar{\wedge}l_1)} + \widetilde{m}_{\Theta(1\bar{\wedge}l_1)} + \widetilde{m}_{2^\Theta(1\bar{\wedge}l_1)})$$
$$= \frac{\prod_{i=1}^{l_1}(m_{\theta i}+m_{\Theta i}+(1-r_i)) - \prod_{i=1}^{l_1}(m_{\Theta i}+(1-r_i))}{k_{1\bar{\wedge}l_1}} + \frac{\prod_{i=1}^{l_1}(m_{\Theta i}+(1-r_i)) - \prod_{i=1}^{l_1}(1-r_i)}{k_{1\bar{\wedge}l_1}} + \frac{\prod_{i=1}^{l_1}(1-r_i)}{k_{1\bar{\wedge}l_1}} \quad \text{(A12-28)}$$
$$= \frac{\prod_{i=1}^{l_1}(m_{\theta i}+m_{\Theta i}+(1-r_i))}{k_{1\bar{\wedge}l_1}}$$

$$(\widetilde{m}_{\Theta(1\bar{\wedge}l_1)} + \widetilde{m}_{2^\Theta(1\bar{\wedge}l_1)})$$
$$= \frac{\prod_{i=1}^{l_1}(m_{\Theta i}+(1-r_i)) - \prod_{i=1}^{l_1}(1-r_i)}{k_{1\bar{\wedge}l_1}} + \frac{\prod_{i=1}^{l_1}(1-r_i)}{k_{1\bar{\wedge}l_1}} = \frac{\prod_{i=1}^{l_1}(m_{\Theta i}+(1-r_i))}{k_{1\bar{\wedge}l_1}} \quad \text{(A12-29)}$$

From Equation (A12-1) to Equation (A12 -3) and the above two equations, we then have

$$\widehat{m}_{\theta(1\bar{\wedge}l)} = (\widetilde{m}_{\theta(1\bar{\wedge}l_1)} + \widetilde{m}_{\Theta(1\bar{\wedge}l_1)} + \widetilde{m}_{2^\Theta(1\bar{\wedge}l_1)})(m_{\theta l}+m_{\Theta l}+(1-r_l))$$
$$-(\widetilde{m}_{\Theta(1\bar{\wedge}l_1)} + \widetilde{m}_{2^\Theta(1\bar{\wedge}l_1)})(m_{\Theta l}+(1-r_l))$$
$$= \frac{\prod_{i=1}^{l_1}(m_{\theta i}+m_{\Theta i}+(1-r_i))}{k_{1\bar{\wedge}l_1}}(m_{\theta l}+m_{\Theta l}+(1-r_l)) - \frac{\prod_{i=1}^{l_1}(m_{\Theta i}+(1-r_i))}{k_{1\bar{\wedge}l_1}}(m_{\Theta l}+(1-r_l)) \quad \text{(A12-30)}$$
$$= \frac{\prod_{i=1}^{l}(m_{\theta i}+m_{\Theta i}+(1-r_i))}{k_{1\bar{\wedge}l_1}} - \frac{\prod_{i=1}^{l}(m_{\Theta i}+(1-r_i))}{k_{1\bar{\wedge}l_1}}$$

$$\widehat{m}_{\Theta(1\bar{\wedge}l)} = (\widetilde{m}_{\Theta(1\bar{\wedge}l_1)} + \widetilde{m}_{2^\Theta(1\bar{\wedge}l_1)})(m_{\Theta l}+(1-r_l)) - \widetilde{m}_{2^\Theta(1\bar{\wedge}l_1)}(1-r_l)$$
$$= \frac{\prod_{i=1}^{l_1}(m_{\Theta i}+(1-r_i))}{k_{1\bar{\wedge}l_1}}(m_{\Theta l}+(1-r_l)) - \frac{\prod_{i=1}^{l_1}(1-r_i)}{k_{1\bar{\wedge}l_1}}(1-r_l) \quad \text{(A12-31)}$$
$$= \frac{\prod_{i=1}^{l}(m_{\Theta i}+(1-r_i))}{k_{1\bar{\wedge}l_1}} - \frac{\prod_{i=1}^{l}(1-r_i)}{k_{1\bar{\wedge}l_1}}$$

$$\widehat{m}_{2^\Theta(1\bar{\wedge}l)} = \widetilde{m}_{2^\Theta(1\bar{\wedge}l_1)}(1-r_l) = \frac{\prod_{i=1}^{l_1}(1-r_i)}{k_{1\bar{\wedge}l_1}}(1-r_l) = \frac{\prod_{i=1}^{l}(1-r_i)}{k_{1\bar{\wedge}l_1}} \quad \text{(A12-32)}$$

Equation (A12-30) to Equation (A12-32) hence hold for any L, from which we then have

$$p(\theta|e_{1\bar{\wedge}l}) = \frac{\widehat{m}_{\theta(1\bar{\wedge}l)}}{\sum_{C\subseteq\Theta}\widehat{m}_{C(1\bar{\wedge}l)}} = \frac{\widehat{m}_{\theta(1\bar{\wedge}l)}}{\sum_{C\in\Theta}\widehat{m}_{C(1\bar{\wedge}l)} + \widehat{m}_{\Theta(1\bar{\wedge}l)}}$$

$$= \frac{\frac{\prod_{i=1}^{l}(m_{\theta i}+m_{\Theta i}+(1-r_i))}{k_{1\bar{\wedge}l_1}} - \frac{\prod_{i=1}^{l}(m_{\Theta i}+(1-r_i))}{k_{1\bar{\wedge}l_1}}}{\sum_{C\in\Theta}\left(\frac{\prod_{i=1}^{l}(m_{Ci}+m_{\Theta i}+(1-r_i))}{k_{1\bar{\wedge}l_1}} - \frac{\prod_{i=1}^{l}(m_{\Theta i}+(1-r_i))}{k_{1\bar{\wedge}l_1}}\right) + \frac{\prod_{i=1}^{l}(m_{\Theta i}+(1-r_i))}{k_{1\bar{\wedge}l_1}} - \frac{\prod_{i=1}^{l}(1-r_i)}{k_{1\bar{\wedge}l_1}}}$$

$$= \frac{\prod_{i=1}^{l}(m_{\theta i}+m_{\Theta i}+(1-r_i)) - \prod_{i=1}^{l}(m_{\Theta i}+(1-r_i))}{\sum_{C\in\Theta}\left(\prod_{i=1}^{l}(m_{Ci}+m_{\Theta i}+(1-r_i)) - \prod_{i=1}^{l}(m_{\Theta i}+(1-r_i))\right) + \prod_{i=1}^{l}(m_{\Theta i}+(1-r_i)) - \prod_{i=1}^{l}(1-r_i)} \tag{A12-33}$$

$$= \frac{\prod_{i=1}^{l}(m_{\theta i}+m_{\Theta i}+(1-r_i)) - \prod_{i=1}^{l}(m_{\Theta i}+(1-r_i))}{\sum_{C\in\Theta}\prod_{i=1}^{l}(m_{Ci}+m_{\Theta i}+(1-r_i)) - N\prod_{i=1}^{l}(m_{\Theta i}+(1-r_i)) + \prod_{i=1}^{l}(m_{\Theta i}+(1-r_i)) - \prod_{i=1}^{l}(1-r_i)}$$

$$= \frac{\prod_{i=1}^{l}(m_{\theta i}+m_{\Theta i}+(1-r_i)) - \prod_{i=1}^{l}(m_{\Theta i}+(1-r_i))}{\sum_{C\in\Theta}\prod_{i=1}^{l}(m_{Ci}+m_{\Theta i}+(1-r_i)) - (N-1)\prod_{i=1}^{l}(m_{\Theta i}+(1-r_i)) - \prod_{i=1}^{l}(1-r_i)}$$

$$p(\Theta|e_{1\bar{\wedge}l}) = \frac{\widehat{m}_{\Theta(1\bar{\wedge}l)}}{\sum_{C\subseteq\Theta}\widehat{m}_{C(1\bar{\wedge}l)}} = \frac{\widehat{m}_{\Theta(1\bar{\wedge}l)}}{\sum_{C\in\Theta}\widehat{m}_{C(1\bar{\wedge}l)} + \widehat{m}_{\Theta(1\bar{\wedge}l)}}$$

$$= \frac{\frac{\prod_{i=1}^{l}(m_{\Theta i}+(1-r_i))}{k_{1\bar{\wedge}l_1}} - \frac{\prod_{i=1}^{l}(1-r_i)}{k_{1\bar{\wedge}l_1}}}{\sum_{C\in\Theta}\left(\frac{\prod_{i=1}^{l}(m_{Ci}+m_{\Theta i}+(1-r_i))}{k_{1\bar{\wedge}l_1}} - \frac{\prod_{i=1}^{l}(m_{\Theta i}+(1-r_i))}{k_{1\bar{\wedge}l_1}}\right) + \frac{\prod_{i=1}^{l}(m_{\Theta i}+(1-r_i))}{k_{1\bar{\wedge}l_1}} - \frac{\prod_{i=1}^{l}(1-r_i)}{k_{1\bar{\wedge}l_1}}} \tag{A12-34}$$

$$= \frac{\prod_{i=1}^{l}(m_{\Theta i}+(1-r_i)) - \prod_{i=1}^{l}(1-r_i)}{\sum_{C\in\Theta}\prod_{i=1}^{l}(m_{Ci}+m_{\Theta i}+(1-r_i)) - (N-1)\prod_{i=1}^{l}(m_{\Theta i}+(1-r_i)) - \prod_{i=1}^{l}(1-r_i)}$$

So, Equation (77) also holds at $L=l$. It therefore holds for any L.

The untrustworthiness for the above results is calculated by the following equation:

$$\widetilde{m}_{2^{\Theta}(1\bar{\wedge}l)} = \frac{\widehat{m}_{2^{\Theta}(1\bar{\wedge}l)}}{\sum_{C\subseteq\Theta}\widehat{m}_{C(1\bar{\wedge}l)} + \widehat{m}_{2^{\Theta}(1\bar{\wedge}l)}} = \frac{\widehat{m}_{2^{\Theta}(1\bar{\wedge}l)}}{\sum_{C\in\Theta}\widehat{m}_{C(1\bar{\wedge}l)} + \widehat{m}_{\Theta(1\bar{\wedge}l)} + \widehat{m}_{2^{\Theta}(1\bar{\wedge}l)}}$$

$$= \frac{\frac{\prod_{i=1}^{l}(1-r_i)}{k_{1\bar{\wedge}l_1}}}{\sum_{C\in\Theta}\left(\frac{\prod_{i=1}^{l}(m_{Ci}+m_{\Theta i}+(1-r_i))}{k_{1\bar{\wedge}l_1}} - \frac{\prod_{i=1}^{l}(m_{\Theta i}+(1-r_i))}{k_{1\bar{\wedge}l_1}}\right) + \frac{\prod_{i=1}^{l}(m_{\Theta i}+(1-r_i))}{k_{1\bar{\wedge}l_1}} - \frac{\prod_{i=1}^{l}(1-r_i)}{k_{1\bar{\wedge}l_1}} + \frac{\prod_{i=1}^{l}(1-r_i)}{k_{1\bar{\wedge}l_1}}} \tag{A12-35}$$

$$= \frac{\prod_{i=1}^{l}(1-r_i)}{\sum_{C\in\Theta}\prod_{i=1}^{l}(m_{Ci}+m_{\Theta i}+(1-r_i)) - (N-1)\prod_{i=1}^{l}(m_{\Theta i}+(1-r_i))}$$

So, Equation (78) also holds at $L=l$. It therefore holds for any L.

A13. Proof of Corollary 2.4 (Bayesian inference)

Given that there is no ambiguity in data and both pieces of evidence are fully reliable, either individually or jointly, and the variables from which the two pieces of evidence are acquired are each the most important, then Equation (59) and Equation (60) reduce to

$$p(\theta|e_{li}\wedge e_{mj}) = \begin{cases} \frac{\widehat{m}_{\theta(li,mj)}}{\sum_{B\in\Theta}\widehat{m}_{B(li,mj)}} & \theta\in\Theta \\ 0 & otherwise \end{cases} \tag{A13-1}$$

$$\widehat{m}_{\theta(li,mj)} = \overline{\omega}_{li,mj}\overline{w}_{\theta(l,m)}\overline{\alpha}_{\theta(li,mj)}m_{\theta li}m_{\theta mj} \qquad \forall\theta\in\Theta \tag{A13-2}$$

From Equation (19), Equation (21), Equation (23), Equation (49), Equation (51), Equation (53) and Equation (54), we have

$$\overline{\omega}_{li,mj} = \frac{\omega_{li,mj}}{\omega_{li}\omega_{mj}} = \frac{\frac{1}{\sum_{\theta\in\Theta} w_{\theta lm} p_{\theta(li,mj)} + 1 - r_{li,mj}}}{\frac{1}{\sum_{\theta\in\Theta} w_{\theta l} p_{\theta li} + 1 - r_{li}} \times \frac{1}{\sum_{\theta\in\Theta} w_{\theta m} p_{\theta mj} + 1 - r_{mj}}}$$

$$= \frac{\left(\sum_{\theta\in\Theta} w_{\theta l} p_{\theta li} + 1 - r_{li}\right) \times \left(\sum_{\theta\in\Theta} w_{\theta m} p_{\theta mj} + 1 - r_{mj}\right)}{\sum_{\theta\in\Theta} w_{\theta lm} p_{\theta(li,mj)} + 1 - r_{li,mj}} \quad \text{(A13-3)}$$

$$= \frac{\left(\sum_{\theta\in\Theta} 1 \times p_{\theta li} + 1 - 1\right) \times \left(\sum_{\theta\in\Theta} 1 \times p_{\theta mj} + 1 - 1\right)}{\sum_{\theta\in\Theta} 1 \times p_{\theta(li,mj)} + 1 - 1}$$

$$= \frac{\left(\sum_{\theta\in\Theta} p_{\theta li}\right) \times \left(\sum_{\theta\in\Theta} p_{\theta mj}\right)}{\sum_{\theta\in\Theta} p_{\theta(li,mj)}} = \frac{1 \times 1}{1} = 1$$

$$\overline{w}_{\theta(l,m)} = w_{\theta lm} / (w_{\theta l} w_{\theta m}) = 1/(1 \times 1) = 1 \quad \text{(A13-4)}$$

$$\overline{\alpha}_{\theta(li,mj)} = \frac{p_{\theta(li,mj)}}{p_{\theta li} p_{\theta mj}} \quad \text{(A13-5)}$$

$$m_{\theta li} = w_{\theta l} p_{\theta li} = p_{\theta li} \text{ and } m_{\theta mj} = w_{\theta m} p_{\theta mj} = p_{\theta mj} \quad \text{(A13-6)}$$

From Equation (48), putting Equation (A13-3) to Equation (A13-6) into Equation (A13-2) gets

$$\widehat{m}_{\theta(li,mj)} = 1 \times 1 \times \frac{p_{\theta(li,mj)}}{p_{\theta li} p_{\theta mj}} \times p_{\theta li} p_{\theta mj} = p_{\theta(li,mj)} = \frac{c_{\theta(li,mj)}}{\sum_{A\in\Theta} c_{A(li,mj)}} \qquad \forall \theta \in \Theta \quad \text{(A13-7)}$$

Putting Equation (A13-7) into Equation (A13-1) for any $\theta \in \Theta$ leads to

$$p(\theta | e_{li} \wedge e_{mj}) = \frac{\widehat{m}_{\theta(li,mj)}}{\sum_{B\in\Theta} \widehat{m}_{B(li,mj)}} = \frac{\frac{c_{\theta(li,mj)}}{\sum_{A\in\Theta} c_{A(li,mj)}}}{\sum_{B\in\Theta} \frac{c_{B(li,mj)}}{\sum_{A\in\Theta} c_{A(li,mj)}}} = \frac{c_{\theta(li,mj)}}{\sum_{B\in\Theta} c_{B(li,mj)}} \qquad \forall \theta \in \Theta \quad \text{(A13-8)}$$

If prior e_0 is acquired independently of both e_{li} and e_{mj}, its combination with both e_{li} and e_{mj}, whose joint basic probability is given by Equation (A13-8), is calculated by Corollary 1.6

$$p(\theta | (e_{li} \wedge e_{mj}) \wedge e_0) = \frac{\sum_{A\cap B=\theta} p(A | e_{li} \wedge e_{mj}) p_{B0}}{1 - \sum_{A\cap B=\phi} p(A | e_{li} \wedge e_{mj}) p_{B0}} = \frac{p(\theta | e_{li} \wedge e_{mj}) p_{\theta 0}}{\sum_{A\in\Theta} p(A | e_{li} \wedge e_{mj}) p_{A0}}$$

$$= \frac{\frac{c_{\theta(li,mj)}}{\sum_{B\in\Theta} c_{B(li,mj)}} \times p_{\theta 0}}{\sum_{A\in\Theta} \frac{c_{A(li,mj)}}{\sum_{B\in\Theta} c_{B(li,mj)}} \times p_{A0}} = \frac{c_{\theta(li,mj)} p_{\theta 0}}{\sum_{A\in\Theta} c_{A(li,mj)} p_{A0}} \qquad \forall \theta \in \Theta \quad \text{(A13-9)}$$

If e_{li} and e_{mj} are also independent of each other, their combination is then calculated by Corollary 1.6 as follows:

$$p(\theta | e_{li} \wedge e_{mj}) = \frac{\sum_{A\cap B=\theta} p_{Ali} p_{Bmj}}{1 - \sum_{A\cap B=\phi} p_{Ali} p_{Bmj}} = \frac{p_{\theta li} p_{\theta mj}}{\sum_{B\in\Theta} p_{Bli} p_{Bmj}} \qquad \forall \theta \in \Theta \quad \text{(A13-10)}$$

If prior e_0 is acquired independently of e_{li} or e_{mj}, its combination with both e_{li} and e_{mj}, whose joint basic probability is given by Equation (A13-10), is calculated by Corollary 1.6

$$p(\theta | (e_{li} \wedge e_{mj}) \wedge e_0) = \frac{\sum_{A\cap B=\theta} p(A | e_{li} \wedge e_{mj}) p_{B0}}{1 - \sum_{A\cap B=\phi} p(A | e_{li} \wedge e_{mj}) p_{B0}} = \frac{p(\theta | e_{li} \wedge e_{mj}) p_{\theta 0}}{\sum_{A\in\Theta} p(A | e_{li} \wedge e_{mj}) p_{A0}}$$

$$= \frac{\frac{p_{\theta li} p_{\theta mj}}{\sum_{B\in\Theta} p_{Bli} p_{Bmj}} \times p_{\theta 0}}{\sum_{A\in\Theta} \frac{p_{Ali} p_{Amj}}{\sum_{B\in\Theta} p_{Bli} p_{Bmj}} \times p_{A0}} = \frac{p_{\theta li} p_{\theta mj} p_{\theta 0}}{\sum_{A\in\Theta} p_{Ali} p_{Amj} p_{A0}} \qquad \forall \theta \in \Theta \quad \text{(A13-11)}$$

From Equation (9), we can finally re-write Equation (A13-11) as follows:

$$p(\theta | (e_{li} \wedge e_{mj}) \wedge e_0) = \frac{\frac{c_{\theta li}}{\sum_{D\in\Theta} c_{Dli}} \times \frac{c_{\theta mj}}{\sum_{D\in\Theta} c_{Dmj}} \times p_{\theta 0}}{\sum_{A\in\Theta} \frac{c_{Ali}}{\sum_{D\subseteq\Theta} c_{Dli}} \times \frac{c_{Amj}}{\sum_{D\in\Theta} c_{Dmj}} \times p_{A0}} = \frac{c_{\theta li}\ c_{\theta mj} p_{\theta 0}}{\sum_{A\in\Theta} c_{Ali}\ c_{Amj} p_{A0}} \quad \forall \theta \in \Theta$$

Appendix B. – Application of the *MAKER* framework to human wellbeing analysis

B1. Problem description and descriptive data analysis

In this section, the *MAKER* framework is applied to human wellbeing analysis based on the British Household Panel Survey (*BHPS*) data [23]. The *BHPS* data collected for this study involves over 10,000 people for a period of 12 years and has 163,043 data records in total [73,75]. The *BHPS* data is used to develop a system model within the *MAKER* framework to analyse the probabilistic relationship between life satisfaction (y) as system output variable and health (x_1) and income (x_2) as system input variables.

In the *BHPS* data, life satisfaction y is categorised into 7 levels, labelled as system states H_n ($n = 1,\cdots,7$) from the lowest satisfaction (H_1) to the highest (H_7), with the system state space defined by $\Theta = \{H_1,\cdots,H_7\}$. In this paper, health x_1 is categorised into 3 grades: G_1 (<0.6), G_2 (0.6-0.8) and G_3 (>0.8). Although more grades are used to categorise health scores from 0 to 1 in the original *BHPS* data [73], the use of 3 grades is appropriate for demonstrating the application of the *MAKER* framework. For the same reason, income x_2 is categorised into 3 bands: B_1 (£0-£19,999), B_2 (£20,000-£39,999) and B_3 (£40,000+).

To demonstrate the application of the *MAKER* rule, we choose to analyse the data from one of the 12 years, named as wave 8, which includes 10,548 data records. Table B-1 and Table B-2 show the one-dimensional contingency tables of the *BHPS* data to represent the recorded frequencies in 7 levels of life satisfaction for 3 health grades and 3 income bands. In Table B-1 and Table B-2, $e_{\theta 1}$ and ($e_{\theta 2}$) is the evidential element that health (income) points to level θ of life satisfaction for $\theta \subseteq \Theta$.

The *BHPS* data is imperfect in that it is ambiguous and imbalanced. In fact, there are 184 cases where there is no answer to the level of life satisfaction, meaning that they could be at any level from H_1 to H_7, or their level of satisfaction is unknown, so these 184 cases are allocated to system space Θ in Table B-1 and Table B-2. The data is imbalanced as understandably people's life satisfaction is not evenly distributed across the spectrum of 7 life satisfaction levels and the number of people is rather different in different health grades and income bands. For example, there are 3403 people highly satisficed at level H_6, compared with only 137 people least satisfied at level H_1, with an imbalance ratio of 25:1; there are 8222 people in the low income band but only 244 people in the high income band, with an imbalance ratio of 34:1.

Table B-1
Contingency Table of Life Satisfaction for Health

Frequency		$e_{H_1 1}$	$e_{H_2 1}$	$e_{H_3 1}$	$e_{H_4 1}$	$e_{H_5 1}$	$e_{H_6 1}$	$e_{H_7 1}$	$e_{\Theta 1}$	Total
Health grade x_1	G_1	56	59	124	147	164	79	73	35	737
	G_2	48	57	162	344	586	495	351	55	2098
	G_3	33	73	277	814	2352	2829	1241	94	7713
Total		137	189	563	1305	3102	3403	1665	184	10548

Table B-2
Contingency Table of Life Satisfaction for Income

Frequency		$e_{H_1 2}$	$e_{H_2 2}$	$e_{H_3 2}$	$e_{H_4 2}$	$e_{H_5 2}$	$e_{H_6 2}$	$e_{H_7 2}$	$e_{\Theta 2}$	Total
Income band x_2	B_1	124	156	467	1091	2355	2470	1399	160	8222
	B_2	11	29	89	193	675	832	230	23	2082
	B_3	2	4	7	21	72	101	36	1	244
Total		137	189	563	1305	3102	3403	1665	184	10548

To analyse the interdependence between health and income in relation to life satisfaction, joint contingency Table B-3 is constructed from the same *BHPS* data, where $e_{\theta 12}$ is the evidential element that both health and income point to level θ of life satisfaction for $\theta \subseteq \Theta$, $S_{\theta(1\overline{\wedge}2)k}$ is the number of people having level θ of life satisfaction and at the k^{th} combination of health grade i and income band j with $k = j + 3(i-1)$, and $S_{(1\overline{\wedge}2)k}$ is the number of people at the k^{th} combination of health grade i and income band j.

Table B-3
Contingency Table of Life Satisfaction for Both Health and Income

Frequency		$e_{H_1 12}$ ($S_{H_1(1\overline{\wedge}2)k}$)	$e_{H_2 12}$ ($S_{H_2(1\overline{\wedge}2)k}$)	$e_{H_3 12}$ ($S_{H_3(1\overline{\wedge}2)k}$)	$e_{H_4 12}$ ($S_{H_4(1\overline{\wedge}2)k}$)	$e_{H_5 12}$ ($S_{H_5(1\overline{\wedge}2)k}$)	$e_{H_6 12}$ ($S_{H_6(1\overline{\wedge}2)k}$)	$e_{H_7 12}$ ($S_{H_7(1\overline{\wedge}2)k}$)	$e_{\Theta 12}$ ($S_{\Theta(1\overline{\wedge}2)k}$)	Total ($S_{(1\overline{\wedge}2)k}$)
G_1	B_1	51	54	115	132	147	71	67	35	672
	B_2	3	3	9	12	14	8	5	0	54
	B_3	2	2	0	3	3	0	1	0	11
G_2	B_1	43	46	135	306	481	381	329	51	1772
	B_2	5	11	25	36	102	106	19	4	308
	B_3	0	0	2	2	3	8	3	0	18
G_3	B_1	30	56	217	653	1727	2018	1003	74	5778
	B_2	3	15	55	145	559	718	206	19	1720
	B_3	0	2	5	16	66	93	32	1	215
Total		137	189	563	1305	3102	3403	1665	184	10548

It can be observed from Table B-3 that the majority of surveyed people are in the lower income bands B_1 and B_2, regardless of their health grades, and among high income band B_3 most people have high health grade G_3, with a rather small group of only 11 people having low health grade G_1. Although this small group of 11 people probably is not sufficient to generate statistically significant conclusions for a combination of high income band and low health grade, the following likelihood analysis and the maximum likelihood inference using the *BHPS* data can still reveal statistically significant patterns for other combinations.

B2. Likelihood analysis of BHPS data

By applying the likelihood method for data analysis [72] as summarised in Section 2.3, normalized likelihoods for health (income) are generated, as shown in Table B-4 (Table B-5), where e_{1i} (e_{2j}) is the evidence acquired from the i^{th} grade of health (the j^{th} band of income) for $i,j = 1,\cdots,3$. Interesting patterns can be observed from these two tables. Suppose middle level H_4 is regarded as neutral, above H_4 as satisfied with life and below H_4 as unsatisfied with life. From Table B-4, people with low health grade G_1 are 7.85 times more likely to be unsatisfied with life than satisfied, with 7.85 equal to 0.69 (0.2997+0.2289+0.1615) divided by 0.0879 (0.0388+ 0.017+0.0321). People with middle health grade G_2 are also 1.72 times more likely to be unsatisfied with life than satisfied. In contrast, people with high health grade G_3 is 2.09 times more likely to be satisfied with life than not.

Similarly, from Table B-5, people in the low income band B_1 are 1.1 times more likely to be unsatisfied than satisfied; however, people in the middle income band B_2 are 1.53 times more likely to be satisfied than not, and people with high income B_3 are 1.55 times more likely to be satisfied than not. Another observation is that people in the lower income band are more likely to be unsure about life satisfaction, i.e. with higher likelihood left for Θ, than other groups of people. Note that these observations are only limited to wave 8 of the *BHPS* data.

Table B-4
Normalized Likelihood of Life Satisfaction for Health

Likelihood		$e_{H_1 1}$	$e_{H_2 1}$	$e_{H_3 1}$	$e_{H_4 1}$	$e_{H_5 1}$	$e_{H_6 1}$	$e_{H_7 1}$	$e_{\Theta 1}$
x_1	$G_1(e_{11})$	0.2997	0.2289	0.1615	0.0826	0.0388	0.0170	0.0321	0.1395
	$G_2(e_{12})$	0.1711	0.1473	0.1405	0.1287	0.0923	0.0710	0.1030	0.1460
	$G_3(e_{13})$	0.0525	0.0842	0.1072	0.1359	0.1652	0.1812	0.1624	0.1113

Table B-5
Normalized Likelihood of Life Satisfaction for Income

Likelihood		$e_{H_1 2}$	$e_{H_2 2}$	$e_{H_3 2}$	$e_{H_4 2}$	$e_{H_5 2}$	$e_{H_6 2}$	$e_{H_7 2}$	$e_{\Theta 2}$
x_2	$B_1(e_{21})$	0.1373	0.1252	0.1259	0.1268	0.1152	0.1101	0.1275	0.1319
	$B_2(e_{22})$	0.0635	0.1213	0.1250	0.1169	0.1720	0.1933	0.1092	0.0988
	$B_3(e_{23})$	0.1012	0.1467	0.0862	0.1116	0.1609	0.2058	0.1499	0.0377

A two-dimensional likelihood table for both health and income is given in Table B-6, where $e_{(1\overline{\wedge}2)k}$ is the evidence acquired from the k^{th} combination of health grade and income band for $k = 1,\cdots,9$. From Table B-6, interesting patterns can also be observed. One of most noticeable patterns is that people in the high income band with low health grade are 16.06 times more likely to be unsatisfied than satisfied, being the most unsatisfied people among all groups, although the sample size of this group is the smallest, with only 11 people. In contrast, people in the high income band with high health grade are 3.48 times more likely to be satisfied than not, being the most satisfised people among all groups. Other patterns can also be observed from Tale B-6.

Table B-6
Normalized Likelihood of Life Satisfaction for Both Health and Income

Likelihood		$e_{H_1(1\overline{\wedge}2)}$	$e_{H_2(1\overline{\wedge}2)}$	$e_{H_3(1\overline{\wedge}2)}$	$e_{H_4(1\overline{\wedge}2)}$	$e_{H_5(1\overline{\wedge}2)}$	$e_{H_6(1\overline{\wedge}2)}$	$e_{H_7(1\overline{\wedge}2)}$	$e_{\Theta(1\overline{\wedge}2)}$
G_1	$B_1(e_{(1\overline{\wedge}2)1})$	0.2950	0.2264	0.1618	0.0801	0.0375	0.0165	0.0319	0.1507
	$B_2(e_{(1\overline{\wedge}2)2})$	0.3007	0.2180	0.2195	0.1263	0.0620	0.0323	0.0412	0.0000
	$B_3(e_{(1\overline{\wedge}2)3})$	0.5026	0.3643	0.0000	0.0791	0.0333	0.0000	0.0207	0.0000
G_2	$B_1(e_{(1\overline{\wedge}2)4})$	0.1770	0.1372	0.1352	0.1322	0.0874	0.0631	0.1114	0.1563
	$B_2(e_{(1\overline{\wedge}2)5})$	0.1383	0.2206	0.1683	0.1045	0.1246	0.1180	0.0432	0.0824
	$B_3(e_{(1\overline{\wedge}2)6})$	0.0000	0.0000	0.3481	0.1502	0.0948	0.2304	0.1766	0.0000
G_3	$B_1(e_{(1\overline{\wedge}2)7})$	0.0616	0.0833	0.1084	0.1407	0.1566	0.1668	0.1694	0.1131
	$B_2(e_{(1\overline{\wedge}2)8})$	0.0236	0.0855	0.1052	0.1197	0.1941	0.2273	0.1333	0.1112
	$B_3(e_{(1\overline{\wedge}2)9})$	0.0000	0.1008	0.0846	0.1168	0.2027	0.2603	0.1831	0.0518

The above observation of interesting patterns may not be 100% reliable as the *BHPS* data is ambiguous and imbalanced. A question is then how trustworthy these patterns can be. To answer this question, we need to investigate how reliable the likelihoods of Table B-4 to Table B-6 may be, which depends on the quality of the *BHPS* data. *BHPS* data collection is a carefully designed and well managed process. It is beyond the scope of this paper to investigate if there are concerns on the data collection process or sample size. However, since the above *BHPS* data is imbalanced, e.g. with a rather small number of people in the high income band, there is a need to investigate whether the *BHPS* data is of appropriate quality for generating the likelihoods of Table B-4 to Table B-6, or how trustworthy it is to use the likelihoods of Table B-6 for inference. To facilitate the investigation, an optimal learning model is built and solved in the next subsection.

B3. Construction of optimal learning model

Theorem 1 can be applied to enhance the above likelihood analysis of the *BHPS* data on the relationships between life satisfaction and two input variables. In this section, we first build an optimal learning model using the data and the above likelihood analysis results, and then use the optimally learnt results to reanalyse the relationships in the next section.

In Equation (59) and Equation (60), interdependence index $\overline{\alpha}_{Ali,Bmj}$ for any $A,\ B \subseteq \Theta$ can be calculated using Equation (49) and the likelihoods of Table B-4 to Table B-6, as shown in Table B-7. The values of $\overline{\alpha}_{Ali,Bmj}$ for $A,\ B \subseteq \Theta$ in Table B-7 are all larger than 1, except for some zero values mostly for the high income band, apparently caused by imbalance and lack of data, indicating that health grades and income bands are highly dependent in people's assessment of life satisfaction. For example, $\overline{\alpha}_{\theta 13,\theta 23} = 7.6217$, 6.9827 and 7.5184 for $\theta = H_5, H_6, H_7$ for the group of people with high health grade G_3 and in high income band B_3. This means that health and income are very highly dependent on each other for this group of people's assessments of life satisfaction, which is consistent with the patterns observed in the previous section.

Table B-7
Interdependence Indices for Every Pair of Health and Income

Interdependence Indices		H_1	H_2	H_3	H_4	H_5	H_6	H_7	Θ
G_1	B_1	7.1669	7.8982	7.9638	7.6505	8.4096	8.8195	7.7803	8.1912
	B_2	15.8084	7.8516	10.8784	13.0781	9.2950	9.8136	11.7479	0.0000
	B_3	16.5693	10.8480	0.0000	8.5895	5.3377	0.0000	4.2910	0.0000
G_2	B_1	7.5315	7.4400	7.6448	8.0966	8.2272	8.0694	8.4887	8.1145
	B_2	12.7331	12.3441	9.5813	6.9451	7.8509	8.5964	3.8460	5.7104
	B_3	0.0000	0.0000	28.7339	10.4552	6.3827	15.7577	11.4392	0.0000
G_3	B_1	8.5436	7.9056	8.0335	8.1622	8.2270	8.3596	8.1820	7.7008
	B_2	7.0799	8.3738	7.8540	7.5316	6.8297	6.4911	7.5140	10.1113
	B_3	0.0000	8.1610	9.1524	7.7006	7.6217	6.9827	7.5184	12.3403

There is no data to calculate $\overline{\alpha}_{Ali,Bmj}$ for $A \neq B$. In Equation (60), joint likelihoods could in theory be taken into account for any level (1-7) of satisfaction in each income band with unknown satisfaction in each health grade, as well as for any level (1-7) of satisfaction in each health grade with unknown satisfaction in each income band. However, the data of Table B-3 is complete in the sense that in any case both health and income values are recorded. This means that any missing value for life satisfaction is associated with both health and income rather than one of them. As such, in the *BHPS* data there is no cross-dependency between any level (1-7) of satisfaction in each income band with unknown satisfaction in each health grade, and vice versa. We then set $\overline{\alpha}_{Ali,Bmj} = 0$ for $A \neq B$ because unknown life satisfaction was not recorded for health or income separately but for both of them jointly.

To infer the level of a person's life satisfaction from her health grade and income band using Equation (59) and Equation (60), weights and reliabilities in the equations need to be given, either assessed subjectively or learnt from data or both. In this section, we show how to use the *BHPS* data to learn these parameters using the optimal learning model of Equation (84) to Equation (96). In this learning model, there are 48 variables in total, including 24 weights $w_{\theta 1}$, $w_{\theta 2}$, and $w_{\theta(1\overline{\wedge}2)}$ for all $\theta \subseteq \Theta$; 15 reliabilities r_{1i}, r_{2j}, and $r_{(1\overline{\wedge}2)k}$ for $i,j = 1, \cdots, 3$ and $k = 1, \cdots, 9$. In addition, since all 10548 observed data records are used for learning, the prior of all the data are also combined to reduce the effect of data imbalance on parameter estimation. The prior distribution of all the data is calculated as follows:

Table B-8
Prior frequency and distribution

	$e_{H_1 0}$	$e_{H_2 0}$	$e_{H_3 0}$	$e_{H_4 0}$	$e_{H_5 0}$	$e_{H_6 0}$	$e_{H_7 0}$	$e_{\Theta 0}$
Frequency	137	189	563	1305	3102	3403	1665	184
e_0	0.0130	0.0179	0.0534	0.1237	0.2941	0.3226	0.1578	0.0174

In Table B-8, $e_{\theta 0}$ is the evidential element that the prior points to life satisfaction level θ for $\theta \subseteq \Theta$. Reliability and weights for the prior are denoted by r_0 and $w_{\theta 0}$ for $\theta \subseteq \Theta$.

A vector of all 48 parameters to be learnt in this case is represented by λ_l as follows:

$$\lambda_l = [w_{\theta 0}, r_0, w_{\theta l}, r_{li}, w_{\theta(1\overline{\wedge}2)}, r_{(1\overline{\wedge}2)k}, \forall \theta \subseteq \Theta, l = 1, 2, i = 1, \cdots, 3, k = 1, \cdots, 9]^T \tag{B-1}$$

The first objective function $f_1(\lambda)$ of the optimal learning model only includes $f_{12}(\lambda)$ as all the 10548 records are complete in that each has both health and income recorded. $p(\theta|e_{(1\overline{\wedge}2)k})$ in Equation (90) is the probability generated by applying Theorem 1 to combine evidence e_{1i} acquired from the i^{th} health grade with evidence e_{2j} acquired from the j^{th} income band with their interdependency taken into account, which is in turn combined with prior e_0 as a piece of independent evidence for all $i,j = 1, \cdots, 3$ and $k = 1, \cdots, 9$. Accordingly, $\widehat{p}_{\theta(A)(1\overline{\wedge}2)k}$ in Equation (90) is the probability of state θ given in a data record for the k^{th} combination of health grade G_i and income band B_j for $k = 1, \cdots, 9$ where state A is observed. From the data of Table B-3, there are $\widehat{p}_{\theta(A)(1\barwedge 2)k} = 1$ at $\theta = A$ and $\widehat{p}_{\theta(A)(1\barwedge 2)k} = 0$ at $\theta \neq A$ for any $\theta, A \subseteq \Theta$ and $k = 1,\cdots,9$; the frequency figures defined in sub-objective function $f_{12}(\lambda)$ are given in Table B-3, with $S_{1\barwedge 2} = S_{(1\barwedge 2)1} + \cdots + S_{(1\barwedge 2)9} = 10548$.

In the second objective function $f_2(\lambda)$, $\widehat{w}_{\theta l}$ in Equation (93) is the reference weight of health ($l = 1$) or income ($l = 2$) for state θ, and is assumed to be the upper bound of $w_{\theta l}$ in this case, or $\widehat{w}_{\theta l} = \overline{w}_{\theta l}$ for $\theta \subseteq \Theta$ under the notion that health and income should play their most important roles in inference for life satisfaction. $\widehat{w}_{\theta(1\barwedge 2)}$ in Equation (95) is the reference weight of both health and income for state θ and is also taken as the upper bound of $\overline{w}_{\theta(1\barwedge 2)}$, or $\widehat{w}_{\theta(1\barwedge 2)} = \overline{w}_{\theta(1\barwedge 2)}$ for any $\theta \subseteq \Theta$. In principle, the reference weights of health and income reflect their perceived importance in assessment of life satisfaction at different health grades or income bands and should be estimated in the data collection processes. $\widehat{r}_{li}$ in Equation (92) is the reference reliability of evidence e_{li}, and is assumed to be the upper bound of r_{li} for any $l = 1, 2,\ i = 1, \cdots, 3$ under the notion that any evidence should be as reliable as possible in inference for life satisfaction. $\widehat{r}_{(1\barwedge 2)k}$ in Equation (94) is the reference reliability of evidence $e_{(1\barwedge 2)k}$ and is assumed as the upper bound of $r_{(1\barwedge 2)k}$ for $k = 1, \cdots, 9$. In principle, the reference reliability of evidence reflects the perceived correctness of the evidence acquired from the collected data and should be estimated in the processes of the data collection and evidence acquisition.

In the optimal learning model, constraints include only upper and lower bounds for the 48 parameters to be learnt. In this study, the data used to acquire evidence is given in Table B-4 for e_{1i} ($i = 1,\cdots,3$), in Table B-5 for e_{2j} ($j = 1,\cdots,3$), and in Table B-6 for $e_{(1\barwedge 2)k}$ ($k = 1,\cdots,9$). These three tables all originate from the same source of the *BHPS* data given in Table B-1 to Table B-3. The *BHPS* data shown in Table B-1 to Table B-3 is complete. Although there is missing data in the *BHPS* data recorded to the state space Θ as unknown, it is recorded separately so that the missing data does not incur cross-interdependency between health grades and unknown for income, and vice versa. However, the *BHPS* data is still imperfect in that it is severely imbalanced, causing concerns on how trustworthy it may be to use the likelihoods of Table B-6 alone for inference.

In fact, such concerns are genuine and can be revealed by applying the rules of thumb, as discussed in Section 2.3. The expected counts for all cells in Table B-3 can be calculated using the simple approach of Row Total times Column Total divided by Total. For instance, the expected count for the cell at the $G_1 - B_1$ row and the $e_{H_1(1\barwedge 2)}$ column is given by $8.7 = \frac{672 \times 137}{10548}$. Table B-9 shows the expected counts for all the cells.

Table B-9
Expected Counts in Each Cell for Joint Health and Income

Expected count		$e_{H_1(1\barwedge 2)}$	$e_{H_2(1\barwedge 2)}$	$e_{H_3(1\barwedge 2)}$	$e_{H_4(1\barwedge 2)}$	$e_{H_5(1\barwedge 2)}$	$e_{H_6(1\barwedge 2)}$	$e_{H_7(1\barwedge 2)}$	$e_{\Theta(1\barwedge 2)}$
G_1	B_1	8.7	12.0	35.9	83.1	197.6	216.8	106.1	11.7
	B_2	0.7	1.0	2.9	6.7	15.9	17.4	8.5	0.9
	B_3	0.1	0.2	0.6	1.4	3.2	3.5	1.7	0.2
G_2	B_1	23.0	31.8	94.6	219.2	521.1	571.7	279.7	30.9
	B_2	4.0	5.5	16.4	38.1	90.6	99.4	48.6	5.4
	B_3	0.2	0.3	1.0	2.2	5.3	5.8	2.8	0.3
G_3	B_1	75.0	103.5	308.4	714.9	1699.2	1864.1	912.1	100.8
	B_2	22.3	30.8	91.8	212.8	505.8	554.9	271.5	30.0
	B_3	2.8	3.9	11.5	26.6	63.2	69.4	33.9	3.8

It is apparent from Table B-9 that the data of Table B-3 does not satisfy the general rule of thumb as there are 22 cells where expected count is < 5 as marked by red and purple ellipses. The data does not satisfy Yates' rule of thumb either as more than 20% of cells with the expected counts being 5 or less have the expected counts of less than 1, as marked by purple ellipses. If these rules are to be followed for statistical inference, questions arise regarding to what extents the joint likelihoods of Table B-6 can be trusted for reasoning and decision making.

The above discussions lead to questioning whether the likelihoods of Table B-4 and Table B-5 should be used to enhance reasoning and decision making. To answer this question, we first calculate the expected counts for all cells for the data of Table B-1 and Table B-2, as given in Table B-10 and Table B-11. The results of Table B-10 show that the data of Table B-1 fully satisfy the general rule of thumb, which means that the evidence acquired for health from Table B-4 should be of good quality for inference. While there are 3 expected counts < 5 in Table B-11, so the general rule of thumb is violated, Yates' rule is satisfied. As such, the evidence acquired for income from Table B-5 should also be of good quality for inference. Although the data of Table B-3 does not satisfy the rules of thumb, it can be used to calculate the interdependence indices in Table B-7. In this context, it is conjectured that combining evidence e_{1i} with evidence e_{2j} for inference should lead to better trusted conclusions.

Table B-10
Expected Counts of Each Cell for Health

Expected count		$e_{H_1 1}$	$e_{H_2 1}$	$e_{H_3 1}$	$e_{H_4 1}$	$e_{H_5 1}$	$e_{H_6 1}$	$e_{H_7 1}$	$e_{\Theta 1}$
x_1	G_1	9.6	13.2	39.3	91.2	216.7	237.8	116.3	12.9
	G_2	27.2	37.6	112.0	259.6	617.0	676.9	331.2	36.6
	G_3	100.2	138.2	411.7	954.3	2268.3	2488.4	1217.5	134.5

Table B-11
Expected Counts of Each Cell for Income

Expected count		$e_{H_1 2}$	$e_{H_2 2}$	$e_{H_3 2}$	$e_{H_4 2}$	$e_{H_5 2}$	$e_{H_6 2}$	$e_{H_7 2}$	$e_{\Theta 2}$
x_2	B_1	106.8	147.3	438.8	1017.2	2418.0	2652.6	1297.8	143.4
	B_2	27.0	37.3	111.1	257.6	612.3	671.7	328.6	36.3
	B_3	3.2	4.4	13.0	30.2	71.8	78.7	38.5	4.3

Following the above analysis, without loss of generality and considering the robust process of collecting the *BHPS* data yet with its observed imbalance, a high confidence of 95% with an error of ±5 in the *BHPS* data is assumed to measure the ability of using the data to generate correct judgements or conclusions. Under this assumption, all lower bounds for weight and reliability are assigned to 0.9, or $\underline{w}_{\theta l} = \underline{r}_{li} = \underline{w}_{\theta(1\overline{\wedge}2)} = \underline{r}_{(1\overline{\wedge}2)k} = 0.9$, and all upper bounds to 1, or $\overline{w}_{\theta l} = \overline{r}_{li} = \overline{w}_{\theta(1\overline{\wedge}2)} = \overline{r}_{(1\overline{\wedge}2)k} = 1$ for all $l = 0,1,2$, $i = 1,\cdots,3$, $\theta \subseteq \Theta$ and $k = 1,\cdots,9$.

The optimal learning model is therefore formulated as a bi-objective nonlinear non-smooth mathematical programming problem, having 2 prioritised nonlinear non-smooth objective functions and 48 variables with their lower and upper bounds as constraints. In this study, Excel Solver is used to solve the problem by selecting the Evolutionary engine.

B4. Results, analysis and discussions

The first objective $f_1(\lambda)$ is of the top priority and is minimised first by setting the starting values of all the parameters to the upper bounds. The optimal solution is given by $f_1^* = f_1(\lambda^*) = 0.0514$. The second objective $f_2(\lambda)$ is then minimised by adding the priority constraint: $f_1(\lambda) \le f_1^* + \delta$ with $\delta = 0.00002$. The optimal solution for the second objective is given by $f_2^* = 0.01$. The detailed solution process using Excel Solver can be found in the provided Supplementary Materials. The optimal results are analysed as follows.

The optimal values of the weights and reliabilities for health, income and the prior are shown in Table B-12, where $w_{\theta l}^*$ is the optimal weight of health (l=1), income (l=2) and the prior (l=0) for state θ. r_{li}^* and t_{li}^* are the optimal reliability and trustworthiness of evidence e_{li} acquired from the i^{th} health grade (l=1) or the i^{th} income band (l=2). r_0^* and t_0^* ($li = 0$) are the optimal reliability and trustworthiness of the prior e_0.

The optimal values of the weights and reliabilities for both health and income are given in Table B-13, where $w_{\theta(1\overline{\wedge}2)^*}$ is the optimal weight of both health and income for state θ. $r_{(1\overline{\wedge}2)k^*}$ and $t_{(1\overline{\wedge}2)k^*}$ are the optimal reliability and trustworthiness of evidence $e_{(1\overline{\wedge}2)k}$ acquired from the k^{th} combination of health grade i and income band j for $i,j = 1,\cdots,3$ and $k = 1,\cdots,9$.

Table B-12
Parameter Values for Health, Income and Prior

Optimal parameter	Health			Income			Prior
	e_{11}	e_{12}	e_{13}	e_{21}	e_{22}	e_{23}	e_0
$w_{H_1 l}^*$	0.9255			0.9067			0.9063
$w_{H_2 l}^*$	0.9432			0.9256			0.9027
$w_{H_3 l}^*$	0.9050			0.9673			0.9005
$w_{H_4 l}^*$	0.9016			0.9212			0.9001
$w_{H_5 l}^*$	0.9976			0.9587			0.9924
$w_{H_6 l}^*$	0.9646			0.9285			0.9961
$w_{H_7 l}^*$	0.9002			0.9190			0.9002
$w_{\Theta l}^*$	0.9081			0.9287			0.9011
r_{li}^*	0.9377	0.9588	0.9996	0.9001	0.9509	0.9627	1.0000
t_{li}^*	0.9368	0.9574	0.9995	0.9031	0.9501	0.9615	1.0000

Table B-13
Parameter Values for Both Health and Income

Optimal parameter	G_1			G_2			G_3		
	B_1	B_2	B_3	B_1	B_2	B_3	B_1	B_2	B_3
	$e_{(1\overline{\wedge}2)1}$	$e_{(1\overline{\wedge}2)2}$	$e_{(1\overline{\wedge}2)3}$	$e_{(1\overline{\wedge}2)4}$	$e_{(1\overline{\wedge}2)5}$	$e_{(1\overline{\wedge}2)6}$	$e_{(1\overline{\wedge}2)7}$	$e_{(1\overline{\wedge}2)8}$	$e_{(1\overline{\wedge}2)9}$
$w_{\theta(1\overline{\wedge}2)^*}$	$w_{H_1(1\overline{\wedge}2)^*} =$ 0.9161, $w_{H_4(1\overline{\wedge}2)^*} =$ 0.9002, $w_{H_7(1\overline{\wedge}2)^*} =$ 0.9001,			$w_{H_2(1\overline{\wedge}2)^*} =$ 0.9040, $w_{H_5(1\overline{\wedge}2)^*} =$ 0.9996, $w_{\Theta(1\overline{\wedge}2)^*} =$ 0.9021			$w_{H_3(1\overline{\wedge}2)^*} =$ 0.9002, $w_{H_6(1\overline{\wedge}2)^*} =$ 0.9692,		
$r_{(1\overline{\wedge}2)k^*}$	0.9579	0.9568	0.9761	0.9729	0.9796	0.9281	0.9071	0.9714	0.9311
$t_{(1\overline{\wedge}2)k^*}$	0.9558	0.9549	0.9745	0.9713	0.9784	0.9279	0.9091	0.9703	0.9317

One observation from the results of Table B-12 and Table B-13 is that the prior is fully reliable and trustworthy with $r_0^* = 1$ and $t_0^* = 1$ but none of the other evidence is fully reliable or trustworthy, i.e. e_{1i} for health, e_{2j} for income and $e_{(1\overline{\wedge}2)k}$ for both health and income for $i,j = 1,\cdots,3$ and $k = 1, \cdots, 9$. This means that the maximum likelihood evidential reasoning is not the same as the straightforward likelihood inference given in Table B-6 for this case, even though e_{1i}, e_{2j} and $e_{(1\bar{\wedge}2)k}$ are all acquired from the same *BHPS* data that is complete and is deemed of high quality. Another observation is that $w_{H_5(1\bar{\wedge}2)^*}$ and $w_{H_6(1\bar{\wedge}2)^*}$ are among the largest of $w_{\theta(1\bar{\wedge}2)^*}$ for all $\theta \subseteq \Theta$ as shown in Table B-13, so are $w^*_{H_5 l}$ and $w^*_{H_6 l}$ among $w^*_{\theta l}$ for all $\theta \subseteq \Theta$ and $l = 1, 2$ as shown in Table B-12. This is due to many more people satisfied at levels H_5 and H_6 than at other levels. The maximum likelihoods $p(\theta|e_{1i} \wedge e_{2j})$ for $i, j = 1, \cdots, 3$ and all $\theta \subseteq \Theta$ are given in Table B-14, which are generated using the optimal weights and reliabilities to combine every pair of health grades and income bands without the prior taken into account.

Comparing the maximum likelihoods of Table B-14 with the ordinary likelihoods of Table B-6, one can find that the formers are larger than the latter for every combination of health grades and income bands for state H_5 and state H_6. This is because optimal weights $w_{H_5(1\bar{\wedge}2)^*}$ and $w_{H_6(1\bar{\wedge}2)^*}$ are among the largest of $w_{\theta(1\bar{\wedge}2)^*}$, and $w^*_{H_5 l}$ and $w^*_{H_6 l}$ among $w^*_{\theta l}$ for all $\theta \subseteq \Theta$ and $l = 1,2$. This is one of the prominent features of *MAKER* in the sense that more important and more reliable evidence is reinforced in the inference process.

Table B-14
Maximum Likelihoods of Life Satisfaction for Joint Health and Income

$p(\theta\|e_{1i} \wedge e_{2j})$		H_1	H_2	H_3	H_4	H_5	H_6	H_7	Θ	$t^*_{1i,2j}$
G_1	$B_1(e_{(1\bar{\wedge}2)1})$	0.2885	0.2202	0.1583	0.0817	0.0453	0.0224	0.0364	0.1473	0.9559
	$B_2(e_{(1\bar{\wedge}2)2})$	0.2877	0.2109	0.2087	0.1219	0.0728	0.0424	0.0440	0.0116	0.9549
	$B_3(e_{(1\bar{\wedge}2)3})$	0.4745	0.3443	0.0104	0.0800	0.0440	0.0122	0.0280	0.0067	0.9746
G_2	$B_1(e_{(1\bar{\wedge}2)4})$	0.1749	0.1362	0.1331	0.1293	0.0965	0.0689	0.1092	0.1521	0.9714
	$B_2(e_{(1\bar{\wedge}2)5})$	0.1358	0.2092	0.1614	0.1034	0.1348	0.1242	0.0472	0.0839	0.9784
	$B_3(e_{(1\bar{\wedge}2)6})$	0.0103	0.0114	0.3211	0.1439	0.1048	0.2336	0.1681	0.0068	0.9280
G_3	$B_1(e_{(1\bar{\wedge}2)7})$	0.0599	0.0815	0.1050	0.1359	0.1693	0.1753	0.1634	0.1097	0.9093
	$B_2(e_{(1\bar{\wedge}2)8})$	0.0245	0.0827	0.1013	0.1159	0.2058	0.2330	0.1295	0.1072	0.9704
	$B_3(e_{(1\bar{\wedge}2)9})$	0.0021	0.0966	0.0820	0.1127	0.2142	0.2655	0.1748	0.0521	0.9318

Another feature of *MAKER* is that it generates optimal values for parameters to measure the extents to which each piece of evidence and their interdependencies should be used as much as possible for more trustworthy inference, whether or not the collected data is of high quality or not. In fact, the maximum likelihoods of Table B-14 are at least as trustworthy as the ordinary likelihoods of Table B-6 because every trustworthiness $t^*_{1i,2j}$ in Table B-14 is no less than the corresponding trustworthiness $t_{(1\bar{\wedge}2)k^*}$ in Table B-13, or $t^*_{1i,2j} \geq t_{(1\bar{\wedge}2)k^*}$ for any $i, j = 1, \cdots, 3$ and $k = 1, \cdots, 9$.

In this case study, the differences between the maximum likelihoods of Table B-14 and the ordinary likelihoods of Table B-6 are not significant. This is due to the data imbalance problem in Table B-3 not being that severe. However, if more health grades or income bands are used for human wellbeing analysis [73,74], the data imbalance problem will become more severe. Also, if more variables than health and income are used for multiple factor analysis of human wellbeing [73,75], the data imbalance problem will become even more pronounced as the number of cells in a joint contingency table increases exponentially with the increase of variables or their discrete referential values or categories. One way to deal with such complex inference problems with limited data is to develop hierarchical models using the belief rule base methodology [68], although this topic is beyond the scope of this paper.

Data availability

Data will be made available on request.

References

[1] F. Almaghrabi, D.L. Xu, J.B. Yang, An evidential reasoning rule based feature selection for improving trauma outcome prediction, Applied Soft Computing 103 (2021) 107112.
[2] N. Balasubramaniam, M. Kauppinen, A. Rannisto, K. Hiekkanen, S. Kujala, Transparency and explainability of AI systems: From ethical guidelines to requirements, Information and Software Technology 159 (2023) 107179.
[3] D. Barber, Bayesian Reasoning and Machine Learning, Cambridge University Press, Cambridge, U.K., 2012.
[4] V. Belton, T. Stewart, Multiple Criteria Decision Analysis: an Integrated Approach, Springer Science & Business Media, 2002.
[5] A. Birnbaum, On the foundations of statistical inference, Journal of the American Statistical Association 57 (298) (1962) 269–306.
[6] I. Campbell, Chi-squared and Fisher–Irwin tests of two-by-two tables with small sample recommendations, Statistics in Medicine 26 (19) (2007) 3661–3675. N.
[7] L.L. Chang, W. Dong, J.B. Yang, X.Y. Sun, X.B. Xu, X.J. Xu, L.M. Zhang, Hybrid belief rule base for regional railway safety assessment with data and knowledge under uncertainty, Information Sciences 518 (2020) 376–395.
[8] Y. Chen, Y.W. Chen, X.B. Xu, C.C. Pan, J.B. Yang, G.K. Yang, A data-driven approximate causal inference model using the evidential reasoning rule, Knowledge-based Systems 88 (2015) 264–272, a.
[9] Y.W. Chen, J.B. Yang, D.L. Xu, S.L. Yang, On the inference and approximation properties of belief rule based systems, Information Sciences 234 (2013) 121–135.
[10] Y.W. Chen, J.B. Yang, C.C. Pan, D.L. Xu, Z.J. Zhou, Identification of uncertain nonlinear systems: Constructing belief rule-based models, Knowledge-Based Systems 73 (2015) 124–133, b.
[11] Y.W. Chen, L. Ni, D.L. Xu, J.B. Yang, Visualising regional disparities in the risk of COVID-19 at different phases of lockdown in England, Environment and Planning A: Economy and Space 53 (5) (2021) 883–886.
[12] A.P. Dempster, Upper and lower probabilities induced by a multi-valued mapping, Annals of Mathematical Statistics 38 (1967) 325–339.

[13] A.P. Dempster, A generalization of Bayesian inference, Journal of the Royal Statistical Society 30 (1968) 205–247.
[14] A.P. Dempster, The Dempster–Shafer calculus for statisticians, International Journal Approximate Reasoning 48 (2) (2008) 365–377.
[15] K. Derrick, X. Liu, Evidential reasoning in the calculation of individual claims reserves, in: IEEE Symposium on Computational Intelligence for Financial Engineering and Economics (IEEE CIFEr), New Jersey, USA, 2024. October.
[16] I.N. Durbach, T.J. Stewart, Modelling uncertainty in multi-criteria decision analysis, European Journal of Operational Research Vol. 223 (1) (2012) 1–14.
[17] R. Fang, H.C. Liao, J.B. Yang, D.L. Xu, Generalised probabilistic linguistic evidential reasoning approach for multi-criteria decision-making under uncertainty, Journal of the Operational Research Society 72 (1) (2021) 130–144.
[18] S. French, Decision theory: An Introduction to the Mathematics of Rationality, Ellis Horwood Limited, 1986.
[19] C. Fu, J.B. Yang, S.L. Yang, A group evidential reasoning approach based on expert reliability, European Journal of Operational Research 246 (2015) 886–893.
[20] C. Fu, M. Xue, W.Y. Liu, D.L. Xu, J.B. Yang, Data-driven preference learning in multiple criteria decision making in the evidential reasoning context, Applied Soft Computing 102 (2021) 107109.
[21] A. Gelman, J.-B. Carlin, H.-S. Stern, D.-B. Dunson, A. Vehtari, D.-B. Rubin, Bayesian Data Analysis, CRC Press, New York, NY, USA, 2014.
[22] C. Howson, P. Urbach, Scientific Reasoning: The Bayesian Approach, 3rd ed., Open Court, Chicago, IL, USA, 2005.
[23] iSER, British Household Panel Survey, Institute for Social and Economic Research, University of Essex, 2021. https://www.iser.essex.ac.uk/bhps.
[24] R.L. Keeney, H. Raiffa, Decisions with Multiple Objectives: Preferences and Value Trade-Offs, Cambridge University Press, 1993.
[25] G.L. Kong, D.L. Xu, J.B. Yang, X.M. Ma, Combined medical quality assessment using the evidential reasoning approach, Expert Systems with Applications 42 (2015) 5522–5530.
[26] G.L. Kong, D.L. Xu, J.B. Yang, X.F. Yin, T.B. Wang, B.G. Jiang, Y.H. Hu, Belief rule-based inference to predict trauma outcome, Knowledge-based Systems 95 (2016) 35–44.
[27] G.L. Kong, L. Jiang, X. Yin, T. Wang, D.L. Xu, J.B. Yang, Y.H. Hu, Combining principal component analysis and the evidential reasoning approach for healthcare quality assessment, Annals of Operations Research 271 (2) (2018) 679–699.
[28] G.L. Kong, D.L. Xu, J.B. Yang, T.B. Wang, B.G. Jiang, Evidential reasoning rule-based decision support system for predicting ICU admission and in-hospital death of trauma, IEEE Transactions on Systems, Man, and Cybernetics: Systems 51 (11) (2021) 7131–7142.
[29] B. Li, H.W. Wang, J.B. Yang, M. Guo, A belief-rule-based inference method for aggregate production planning under uncertainty, International Journal of Production Research 51 (1) (2013) 83–105.
[30] Y.Z. Li, Q.H. Wu, L. Jiang, J.B. Yang, D.L. Xu, Optimal power system dispatch with wind power integrated using nonlinear interval optimization and evidential reasoning approach, IEEE Transactions on Power Systems 31 (3) (2016) 2246–2254.
[31] H.C. Liao, R. Fang, J.B. Yang, D.L. Xu, A linguistic belief-based evidential reasoning approach and its application in aiding lung cancer diagnosis, Knowledge-Based Systems 253 (2022) 109559.
[32] F. Liu, W.D. Zhu, Y.W. Chen, D.L. Xu, J.B. Yang, Evaluation, ranking and selection of R&D projects by multiple experts: an evidential reasoning rule-based approach, Scientometrics 111 (3) (2017) 1501–1519.
[33] F. Liu, Y.W. Chen, J.B. Yang, D.L. Xu, W.S. Liu, Solving multiple criteria R&D project selection problems with a data-driven evidential reasoning rule, International Journal of Project Management 37 (1) (2019) 87–97, a.
[34] J.P. Liu, X.W. Liao, J.B. Yang, A group decision-making approach based on evidential reasoning for multiple criteria sorting problem with uncertainty, European Journal of Operational Research 246 (3) (2015) 858–873.
[35] X. Liu, S. Sachan, J.B. Yang, D.L. Xu, Maximum likelihood evidential reasoning-based hierarchical inference with incomplete data, in: 25th IEEE International Conference on Automation and Computing (ICAC 2019), Lancaster, UK, 2019 b.
[36] X. Liu, J.B. Yang, D.L. Xu, K. Derrick, C. Stubbs, M. Stockdale, Automobile insurance fraud detection using the evidential reasoning approach and data-driven inferential modelling, in: 2020 IEEE International Conference on Fuzzy Systems (FUZZ 2020), Glasgow, UK, 2020.
[37] A.F. Markus, J.A. Kors, P.R. Rijnbeek, The role of explainability in creating trustworthy artificial intelligence for health care: A comprehensive survey of the terminology, design choices, and evaluation strategies, Journal of Biomedical Informatics 113 (2021) 103655.
[38] J. Min, J.J. Zhu, J.B. Yang, The risk monitoring of the financial ecological environment in Chinese outward foreign direct investment based on a complex network, Sustainability 12 (22) (2020) 9456.
[39] B. Mittelstadt, Interpretability and transparency in artificial intelligence, in: Carissa Véliz (Ed.), The Oxford Handbook of Digital Ethics, online edn, Oxford Academic, 2021, https://doi.org/10.1093/oxfordhb/9780198857815.013.20, 10 Nov.
[40] Mos & Co Solicitors, "Jury trial – how your fate is decided", 2017. https://www.mosslaw.co.uk/jury-trial-fate-decided/).
[41] L.A. Ribeiro, A.C. Garcia, P.S. dos Santos, Dependency factors in evidence theory: an analysis in an information fusion scenario applied in adverse drug reactions, Sensors 22 (6) (2022) 2310.
[42] L. Rivadeneira, J.B. Yang, M. López-Ibáñez, Predicting tweet impact using a novel evidential reasoning prediction method, Expert Systems with Applications 169 (2021) 114400.
[43] S. Sachan, J.B. Yang, D.L. Xu, D. Erasobenavides, Y. Li, An explainable AI decision-support-system to automate loan underwriting, Expert Systems With Applications 144 (2020) 1–24.
[44] S. Sachan, F. Almaghrabi, J.B. Yang, D.L. Xu, Evidential reasoning for pre-processing uncertain categorical data for trustworthy decisions: an application on healthcare and finance, Expert Systems with Applications 185 (2021) 115597 a.
[45] S. Sachan, J.B. Yang, D.L. Xu, K. Derrick, X. Liu, C. Stubbs, H. Maddocks, M. Chong, Augmented intelligence for transparent decision making in insurance claims, in: 31st European Conference on Operational Research, Athens, Greece, 2021 b.
[46] G. Shafer, The Mathematical Theory of Evidence, Princeton University Press, 1976.
[47] T.J. Stewart, A critical survey on the status of multiple criteria decision making theory and practice, Omega 20 (1992) 569–586.
[48] P. Sureeyatanapas, J.B. Yang, D. Bamford, Evaluation of corporate sustainability, Frontiers of Engineering Management 1 (2) (2014) 176–194.
[49] P. Sureeyatanapas, J.B. Yang, D. Bamford, The sweet spot in sustainability: a framework for corporate assessment in sugar manufacturing, Production Planning and Control 26 (13) (2015) 1128–1144.
[50] P. Sureeyatanapas, J.B. Yang, D. Bamford, Analysis of criteria weights for the assessment of corporate sustainability: a case study in sugar manufacturing, KKU Engineering Journal 43 (3) (2016) 135–145.
[51] M.M. Tan, D.L. Xu, J.B. Yang, Corporate Failure Risk Assessment for Knowledge-Intensive Services Using the Evidential Reasoning Approach, Journal of Risk and Financial Management 15 (3) (2022) 131.
[52] S.W. Tang, Z.J. Zhou, C.H. Hu, J.B. Yang, Y. Cao, Perturbation analysis of evidential reasoning rule, IEEE Transactions on Systems, Man, and Cybernetics: Systems 51 (8) (2021) 4895–4910.
[53] A. Taroun, J.B. Yang, A DST-based approach for construction project risk analysis, Journal of the Operational Research Society 64 (8) (2013) 1221–1230.
[54] I.B. Torre, A. Riba, J.B. Yang, Analytical tool adoption level: A case study based on an evidential reasoning approach, International Journal of Transitions and Innovation Systems 4 (1/2) (2015) 22–42.
[55] G. Wang, H.R. Wang, Y. Yang, D.L. Xu, J.B. Yang, F. Yue, Group article recommendation based on ER rule in scientific social networks, Applied Soft Computing 110 (2021) 107631.
[56] Y.M. Wang, J.B. Yang, D.L. Xu, Environmental impact assessment using the evidential reasoning approach, European Journal of Operational Research 174 (3) (2006) 1885–1913.
[57] T. Wu, D.L. Xu, J.B. Yang, Decentralised energy and its performance assessment models, Frontiers of Engineering Management 8 (2) (2021) 183–198.
[58] D.L. Xu, J.B. Yang, Y.M. Wang, The ER approach for multi-attribute decision analysis under interval uncertainties, European Journal of Operational Research 174 (3) (2006) 1914–1943.
[59] D.L. Xu, J. Liu, J.B. Yang, G.P. Liu, J. Wang, I. Jenkinson, J. Ren, Inference and learning methodology of belief-rule-based expert system for pipeline leak detection, Expert Systems with Applications 32 (1) (2007) 103–113.

[60] D.L. Xu, C. Foster, Y. Hu, J.B. Yang, Decision support system for evaluating impact of product carbon labelling scheme, Frontiers of Engineering Management 1 (2014) 1–16.
[61] X.B. Xu, J. Zheng, D.L. Xu, J.B. Yang, Information fusion method for fault diagnosis based on evidential reasoning rule, Control Theory & Applications 32 (9) (2015) 1170–1182.
[62] X.B. Xu, J. Zheng, J.B. Yang, D.L. Xu, Y.W. Chen, Data classification using evidence reasoning rule, Knowledge-Based Systems 116 (2017) 144–151.
[63] X.J. Xu, X.P. Yan, C.X. Sheng, C.Q. Yuan, D.L. Xu, J.B. Yang, A belief rule based expert system for fault diagnosis of marine diesel engines, IEEE Transactions on Systems, Man, and Cybernetics: Systems 50 (2) (2020) 656–672.
[64] J.B. Yang, M.G. Singh, An evidential reasoning approach for multiple attribute decision making with uncertainty, IEEE Transactions on Systems, Man, and Cybernetics 24 (1) (1994) 1–18.
[65] J.B. Yang, Minimax reference point approach and its application for multiobjective optimisation, European Journal of Operational Research 126 (2000) 541–556.
[66] J.B. Yang, Rule and utility based evidential reasoning approach for multiple attribute decision analysis under uncertainty, European Journal of Operational Research 131 (1) (2001) 31–61.
[67] J.B. Yang, D.L. Xu, On the evidential reasoning algorithm for multiattribute decision analysis under uncertainty, IEEE Transactions on Systems, Man, and Cybernetics Part A: Systems and Humans 32 (3) (2002) 289–304.
[68] J.B. Yang, Y.M. Wang, D.L. Xu, K.S. Chin, The evidential reasoning approach for MCDA under both probabilistic and fuzzy uncertainties, European Journal of Operational Research 171 (1) (2006) 309–343, a.
[69] J.B. Yang, J. Liu, J. Wang, H.S. Sii, H.W. Wang, A belief rule-base inference methodology using the evidential reasoning approach – RIMER, IEEE Transactions on Systems, Man, and Cybernetics – Part A: Systems and Humans 36 (2) (2006) 266–285, b.
[70] J.B. Yang, J. Liu, D.L. Xu, J. Wang, H.W. Wang, Optimization models for training belief rule based systems, IEEE Transactions on Systems, Man, and Cybernetics – Part A 37 (4) (2007) 569–585.
[71] J.B. Yang, D.L. Xu, Evidential reasoning rule for evidence combination, Artificial Intelligence 205 (2013) 1–29.
[72] J.B. Yang, D.L. Xu, X.B. Xu, C. Fu, Likelihood analysis of imperfect data" and "Supplemental Materials, IEEE Transactions on Systems, Man and Cybernetics: Systems 58 (8) (2023) 5046–5057.
[73] L. Yang, Subjective and Preference-Sensitive Multidimensional Well-Being and Inequality, PhD Thesis of, London School of Economics and Political Science, 2017.
[74] L. Yang, Measuring well-being: a multidimensional index integrating subjective well-being and preferences, Journal of Human Development and Capabilities 19 (4) (2018) 456–476.
[75] Y. Yang, D.L. Xu, J.B. Yang, Y.W. Chen, An evidential reasoning-based decision support system for handling customer complaints in mobile telecommunications, Knowledge-Based Systems 162 (2018) 202–210.
[76] Y. Yang, R.X. Lu, M. Xue, Z.Q. Shou, J.B. Yang, L. Fu, Data-driven evidential reasoning method for evaluating e-government performance, International Journal of Information Technology & Decision Making 20 (1) (2021) 261–285.
[77] S.Y. Yao, J.B. Yang, D.L. Xu, P. Dark, Probabilistic modelling approach for interpretable inference and prediction with data for sepsis diagnosis, Expert Systems with Applications 183 (2021) 1–15, a.
[78] S.Y. Yao, J.B. Yang, D.L. Xu, A probabilistic modelling approach for interpretable data inference and classification, Journal of Intelligent and Fuzzy Systems 40 (3) (2021) 5101–5117, b.
[79] S.Y. Yao, H. Hu, J.B. Yang, A probabilistic safety assessment framework for arctic marine traffic using the evidential reasoning approach, International Journal of Shipping and Transport Logistics 14 (3) (2022) 266–301.
[80] D. Yates, D. Moore, G. McCabe, The Practice of Statistics, Freeman, New York, 1999.
[81] M. Zhou, X.B. Liu, J.B. Yang, C. Fang, Group evidential reasoning approach for MADA under fuzziness and uncertainties, International Journal of Computational Intelligence Systems 6 (3) (2013) 423–441, a.
[82] M. Zhou, X.B. Liu, J.B. Yang, Evidential reasoning approach for MADM based on incomplete interval value, Journal of Intelligent and Fuzzy Systems 33 (6) (2017) 3707–3721.
[83] M. Zhou, X.B. Liu, J.B. Yang, Y.W. Chen, J. Wu, Evidential reasoning approach with multiple kinds of attributes and entropy-based weight assignment, Knowledge-Based Systems 163 (2019) 358–375.
[84] M. Zhou, X.B. Liu, Y.W. Chen, X.F. Qian, J.B. Yang, J. Wu, Assignment of attribute weights with belief distributions for MADM under uncertainties, Knowledge-Based Systems 189 (2020) 105110.
[85] M. Zhou, Y.K. Qiao, J.B. Yang, Y.J. Zhou, X.B. Liu, J. Wu, A three-level consensus model for large-scale multi-attribute group decision analysis based on distributed preference relations under social network analysis, Expert Systems with Applications 204 (2022) 117603.
[86] W.D. Zhu, F. Liu, Y.W. Chen, J.B. Yang, D.L. Xu, D.P. Wang, Research project evaluation and selection: an evidential reasoning rule-based method for aggregating peer review information with reliabilities, Scientometrics 105 (2015) 1469–1490, a.
[87] Z.G. Zhou, F. Liu, L.C. Jiao, Z.J. Zhou, J.B. Yang, M.G. Gong, X.P. Zhang, A bi-level belief rule based decision support system for diagnosis of lymph node metastasis in gastric cancer, Knowledge-Based Systems 54 (2013) 128–136, b.
[88] Z.G. Zhou, F. Liu, L.L. Li, L.C. Jiao, Z.J. Zhou, J.B. Yang, Z.L. Wang, A cooperative belief rule based decision support system for lymph node metastasis diagnosis in gastric cancer, Knowledge-based Systems 85 (2015) 62–70, b.